A STUDY OF WAVELET PACKET FRAMES AND DIFFUSION FILTERS

THESIS

SUBMITTED FOR THE AWARD OF THE DEGREE OF

Doctor of Philosophy

IN

MATHEMATICS

By

JAVID IQBAL

Under the Supervision of

Dr. Md. Kalimuddin Ahmad

DEPARTMENT OF MATHEMATICS
ALIGARH MUSLIM UNIVERSITY
ALIGARH - 202002 (INDIA)

January, 2008

This thesis is dedicated to
my beloved parents
whose endless blessings have
made this thesis possible

Dr. M.K. Ahmad
M.Phil., M.Sc. Nat. (Germany), Ph.D.
DAAD Fellow (Germany) 1991-93 & 1996-98
Reader

Department of Mathematics
Aligarh Muslim University, Aligarh
U.P. - 202 002 (INDIA)
✆ : +91(0)571-2704736 (R)
09837000608 (M)
Fax : +91(0) 571-2704229

CERTIFICATE

This is to certify that the material presented in the thesis entitled "***A Study of Wavelet Packet Frames and Diffusion Filters*** " is the research work of ***Mr. Javid Iqbal*** carried out under my supervision. This work is more than adequate for the partial fulfillment for the award of the degree of Doctor of Philosophy in Mathematics.

I further certify that the work has not been submitted either partly or fully to any other University or Institution for the award of any degree.

MKAhmad

Dr. M. K. Ahmad

(Supervisor)

CHAIRMAN
DEPARTMENT OF MATHEMATICS
A.M.U., ALIGARH 27/10/09

Specializations : Image Processing, Wavelets, Mathematical Modeling and V. Inequalities
Residence : T-1, Azim Estate, New Sir Syed Nagar, Aligarh U.P.
E-mails : ahmad_kalimuddin@yahoo.co.in; mmt13mka@amu.ac.in

# Acknowledgment

At the outset, I bow before Almighty, the Lord of the world, who has taught the use of pen and whose benign benediction granted me the courage, patience and strength to embark upon this work and carry it to its successful completion.

It is good fortune and a matter of pride and privilege for me to have the esteemed supervision of **Dr. Md. Kalimuddin Ahmad**, Reader, Department of Mathematics, Aligarh Muslim University, Aligarh, whose rigorous training and constant encouragement were solely instrumental in getting the work completed. I wish to express my sincere thanks and deep sense of gratitude for his most readily accessible, intelligent, invaluable suggestion, discussion and scholarly guidance, which he offered to me in the midst of his busy schedule. I wish providential to get a life time opportunity to work under his knowledgeable supervision. I take this opportunity to put on record my profound indebtedness to him.

I extend my sincere thanks to **Prof. Huzoor H. Khan**, Chairman, Department of Mathematics, Aligarh Muslim University, Aligarh for providing me all facilities available in the department in carrying out this work and encouragement in pursuit of course of study.

My deep sense of gratitude to all my teachers for their generous help and cooperation, which in turn helped me towards perfection.

It is my pleasure to express my heartiest appreciation to all my friends and colleagues, who have been actively involved in offering suggestion or reviewing and commenting on my thesis from time to time. I am also thankfull to my friends at IIT's and other institutions for their fruitfull advices besides providing copies of various papers not available in Aligarh whenever, I approached them.

I also thankfull to the staff of the Department of Mathematics, A. M. U., Aligarh especially seminar staff and office staff for their nice co-operative attitude.

At this juncture, I do not want to fail in remembering my brothers, sisters, relatives and all my near and dear ones for their deep affection, constant encouragement, loving advise, moral support and as a mentor shaped my life to this stage.

I feel a deep sense of gratitude to my parents, who have been dreaming for quite a long time for my success. They have done everything possible to see me at this place. They have always helped me in good and bad times alike to keep me focused

towards my goal. The completion of this thesis would have been rather impossible without their wholehearted support, encouragement and sacrifices during the tough and easier phases of this project. I feel shortage of words to pay thanks to them.

Finally, I express my indebtedness to my glorious and esteemed institution, Aligarh Muslim University, Aligarh and my homage to its founder Sir Syed Ahmad Khan.

Dated: 22-01-08 **(Javid Iqbal)**

Preface

The theory of Wavelets and diffusion filters are two comparatively new developments of scientific knowledge. Coincidentally both were formally invented in early eighties, but their traces are available in literature since long. Mathematicians knew their examples and properties but were unable to exploit the potentiality of their properties to understand real world problems. In short, wavelet methods are refinement of Fourier analysis, invented in 1984, whose mathematical foundation was provided by Grossman, Morlet, Meyer, Mallat, Daubechies, Coiffman and Wickerhauser et al.

The wavelet theory provides a unified framework for a number of techniques which had been developed independently for various applications. For example, multiresolution signal processing, used in computer vision; subband coding, developed for speech and image compression; and wavelet series expansions, developed in applied mathematics, have been recently recognized as different views of a single theory.

Introduced by Duffin and Schaeffer [32] in the context of non-harmonic Fourier series, the theory of frames has been developed for Gabor and Wavelets by many authors, see especially the papers by Daubechies [24], Heil and Walnut [45], Christensen [14], Sun and Zhou [66], Shang and Zhou [64] and Yang and Zhou [78]. It has application in finding numerically stable algorithm for reconstruction of signals from its atomic decomposition.

On the other hand, the PDE base diffusion, starting with Perona-Malik [62] and extended by Weickert and many others [11, 59, 72, 73], has attracted the attention of Mathematicians and people working in image processing domain. It is a technique for image enhancement / restoration/ edge detection. Since long, wavelet shrinkage was independently used for the same purpose, but recently a beautiful connection has been established between wavelet shrinkage and nonlinear diffusion [72].

The present thesis contains five chapters. In Chapter 1, we have given an overview of wavelets, wavelet packets, theory of frames, some examples, definitions and theorems which are used in the subsequent chapters. In Chapter 2, we have generalised various results of wavelets for wavelet packet frames and obtained their frame bounds in different settings.

The Chapter 3 deals with the concept of a new system of Coherent States, the "*vector valued Weyl-Heisenberg wavelets*", since the group considered here is the Weyl-Heisenberg with a dilation parameter. In sub-sections, 3.2 and 3.3, we have described some notations and the vector valued multiresolution analysis. In subsection 3.4, we have discussed the Weyl-Heisenberg wavelets. In the last section, we have presented the generic construction of frames and proved a result related to the corresponding frame bounds.

In Chapter 4, we focus mainly on the techniques of diffusion based on partial differential equations (PDE's). We have discussed the detailed account of many different diffusion filter models especially by Perona-Malik and Weickert. The discretization scheme and the computational results are also given. In the last section, we have given the wavelet shrinkage scheme and the correspondence between diffusion and shrinkage functions.

In Chapter 5, we have introduced a new diffusion filter based on Sharp Operator. We studied the Hardy-Littlewood maximal function and the sharp operator to measure the oscillatory behaviour of images. With the sharp operator we have measured the oscillation in the neighbourhood of each pixel of an image. We have proved the existence theorem for our diffusion model and presented some computational results.

In the end, a comprehensive bibliography is included.

Contents

Chapter 1

Preliminaries

1.1 Introduction

The fundamental idea behind wavelets is to analyze according to scale. Indeed, some researchers in the wavelet field feel that by using wavelets, one is adopting a whole new mindset or perspective in processing data.

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. This idea is not new. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions. However, in wavelet analysis, the scale that we use to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large "window", we would notice gross features. Similarly, if we look at a signal with a small "window", we would notice small features. The results in wavelet analysis is to see both the forest and the trees, so to speak.

This makes wavelets interesting and useful. For many decades, Scientists have wanted more appropriate functions than the sines and cosines which comprise the bases of Fourier analysis, to approximate choppy signals [22]. By their definition, these functions are non-local (and stretch out to infinity). They therefore do a very poor job in approximating sharp spikes. But with wavelet analysis, we can use approximating functions that are contained neatly infinite domains. Wavelets are well-suited for approximating data with sharp discontinuities.

The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet function), data operations can be performed using just the corresponding wavelet coefficients. And if you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression.

Other applied fields that are making use of wavelets include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, and pure mathematics applications such as solving partial differential equations.

In this chapter, we have given the relevent definitions and results which are used in the subsequent chapters.

1.2 An overview of wavelet

In the history of mathematics, wavelet analysis shows many different origins [58]. Much of the work was performed in the 1930s, and, at the time, the separate efforts did not appear to be part of a coherent theory.

Before 1930, the main branch of mathematics leading to wavelets began with Joseph Fourier (1807) with his theories of frequency analysis, now often referred to as Fourier synthesis. He asserted that any 2π-periodic function $f(x)$ is the sum

$$a_0 + \sum_{k=1}^{\infty}(a_k coskx + b_k sinkx)$$

of its Fourier series. The coefficients a_0, a_k and b_k are calculated by

$$a_0 = \frac{1}{2\pi}\int_0^{\infty} f(x)dx,$$
$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)cos(kx)dx,$$
$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)sin(kx)dx.$$

Fourier's assertion played an essential role in the evolution of the ideas mathematicians had about the functions. He opened up the door to a new functional universe.

After 1807, by exploring the meaning of functions, Fourier series convergence, and orthogonal systems, mathematicians gradually were led from their previous notion of frequency analysis to the notion of scale analysis. That is, analyzing $f(x)$ by creating mathematical structures that vary in scale. How? Construct a function, shift it by some amount, and change its scale. Apply that structure in approximating a signal. Now repeat the procedure. Take that basic structure, shift it, and scale it again. Apply it to the same signal to get a new approximation. And so on. It turns out that this sort of scale analysis is less sensitive to noise because it measures the average fluctuations of the signal at different scales.

The first mention of wavelets appeared in an appendix to the thesis of A. Haar (1909). One property of the Haar wavelet is that it has compact support, which means that it vanishes outside of a finite interval. Unfortunately, Haar wavelets are not continuously differentiable which somewhat limits their applications.

In the 1930s, several groups working independently researched the representation of functions using scale-varying basis functions. Understanding the concepts of basis functions and scale-varying basis functions is key to understanding wavelets.

By using a scale-varying basis function called the Haar basis function, Paul Levy, a 1930s physicist, investigated Brownian motion, a type of random signal [58]. He found the Haar basis function superior to the Fourier basis functions for studying small complicated details in the Brownian motion.

Another 1930s research effort by Littlewood, Paley, and Stein involved computing the energy of a function $f(x)$:

$$\text{energy} = \frac{1}{2}\int_0^{2\pi} |f(x)|^2 dx.$$

The computation produced different results if the energy was concentrated around a few points or distributed over a large interval. This result disturbed the scientists because it indicated that energy might not be conserved. The researchers discovered a function that can vary in scale and can conserve energy when computing the functional energy. Their work provided David Marr with an effective algorithm for numerical image processing using wavelets in early 1980s.

Between 1960 and 1980, the mathematicians Guido Weiss and Ronald R. Coifman studied the simplest elements of a function space, called atoms, with the goal of finding the atoms for a common function space using these atoms. In 1980, Grossman and Morlet, a physicist and an engineer, broadly defined wavelets in the context of quantum physics. These two researchers provided a way of thinking for wavelets based on physical intuition.

In 1985, Stephane Mallat gave wavelets an additional jump-start through his work in digital signal processing. He studied some relationships between quadrature mirror filters, pyramid algorithms, and orthonormal wavelet bases. Inspired in part by three results, Y. Meyer constructed the first non-trivial wavelets. Unlike the Haar wavelets, the Meyer wavelets are continuously differentiable, however they do not have compact support. A couple of years later, Ingrid Daubechies used Mallat's work to construct a set of wavelet orthonormal basis functions that are perhaps the most elegant, and have become the cornerstone of wavelet applications today.

1.3 Fourier analysis/ Fourier transform

Fourier's representation of functions as a superposition of sines and cosines has become ubiquitous for both the analytic and numerical solution of differential equation and for the analysis and treatment of communication signals. Fourier and wavelet analysis have some very strong links.

The Fourier transform utility lies in its ability to analyze a signal in the time domain for its frequency content. The transform works by first translating a function in the time domain into a function in the frequency domain. The signal can then be analyzed for its frequency content because the Fourier coefficients of the transformed function represent the contribution of each sine and cosine function at each frequency. The inverse Fourier transform, transforms data from the frequency domain into the time domain.

Definition 1.3.1 Let Ω be a domain in $\mathbb{R}^n$ and let p be a positive real number. We denote $L^p(\Omega)$ the class of all measurable functions f, defined on Ω for which

$$\int_\Omega |f(x)|^p \, dx < \infty,$$

and the norm of $f \in L^p(\Omega)$ is defined by

$$\|f\|_p := \begin{cases} \left\{ \int_\Omega |u(x)|^p dx \right\}^{1/p}, & \text{for } 1 \le p < \infty \\ \operatorname{ess} \sup\limits_{-\infty < x < \infty} |f(x)|, & \text{for } p = \infty. \end{cases}$$

Definition 1.3.2 We may define the inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)} \, dx, \qquad f, g \in L^2(\mathbb{R}).$$

Definition 1.3.3 The Fourier transform of a function $f \in L^1(\mathbb{R})$ is defined by

$$\hat{f}(w) = (\mathcal{F}f)(w) := \int_{-\infty}^{\infty} e^{-iwx} f(x)\, dx.$$

Definition 1.3.4 Let $\hat{f} \in L^1(\mathbb{R})$ be the Fourier transform of some function $f \in L^1(\mathbb{R})$ then the inverse Fourier transform of $\hat{f}$ is defined by

$$(\mathcal{F}^{-1}\hat{f})(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixw} \hat{f}(w)\, dw.$$

Theorem 1.3.5 Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ then the Fourier transform $\hat{f}$ of f is in $L^2(\mathbb{R})$ and satisfies the following "Parseval identity"

$$\|\hat{f}\|_2^2 = 2\pi \|f\|_2^2.$$

Definition 1.3.6 Let f and g be functions in $L^1(\mathbb{R})$ then (continuous-time) convolution of f and g is also an $L^1(\mathbb{R})$ function h, defined by

$$h(x) = (f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\, dy.$$

1.4 Windowed Fourier transform

If $f(t)$ is a nonperiodic signal, the summation of the periodic functions, sine and cosine, does not accurately represent the signal. One could artificially extend the signal to make it periodic but it would require additional continuity at the endpoints. The windowed Fourier transform (WFT) is one solution to the problem of better representing the nonperiodic signal. The WFT can be used to give information about signals simultaneously in the time domain and in the frequency domain.

With the WFT, the input signal $f(t)$ is chopped up into sections, and each section is analyzed for its frequency content separately. If the signal has sharp transitions, we window the input data so that the sections converge to zero at the endpoints [50]. This windowing is accomplished via a weight function that places less emphasis near the intervals endpoints than in the middle. The effect of the window is to localize the signal in time.

To approximate a function by samples, and to approximate the Fourier integral by the discrete Fourier transform, requires applying a matrix whose order is the number sample points n. Since multiplying an $n \times n$ matrix by a vector costs on the order of n^2 arithmetic operations, the problem gets quickly worse as the number of sample points increases. However, if the samples are uniformly spaced, then the Fourier matrix can be factored into a product of just a few sparse matrices, and the resulting factors can be applied to a vector in a total of order $n\log n$ arithmetic operations. This is the so-called fast Fourier transform (FFT) [63].

Definition 1.4.1 For any fixed value of $\alpha > 0$, the "Gabor Transform or windowed Fourier transform" of an $f \in L^2(\mathbb{R})$ is defined by

$$(\mathcal{G}_b^\alpha f)(w) := \int_{-\infty}^{\infty} (e^{-iwt} f(t)) g_\alpha(t-b)\, dt,$$

where $g_\alpha(t-b)$ is a Gaussian function.

Definition 1.4.2 A nontrivial function $w \in L^2(\mathbb{R})$ is called a window function if $xw(x)$ is also in $L^2(\mathbb{R})$. The centre t^* and radius Δ_w of a window function w are defined as:

$$t^* := \frac{1}{\|w\|_2^2} \int_{-\infty}^{\infty} x|w(x)|^2 dx,$$

and

$$\Delta_w := \frac{1}{\|w\|_2} \Big\{ \int_{-\infty}^{\infty} (x-t^*)^2 |w(x)|^2 dx \Big\}^{1/2},$$

respectively, and the width of the window function w is defined by $2\Delta_w$.

Gabor used "Gaussian function" as window function for time localization and defined by

$$g_\alpha(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{-t^2/4\alpha}, \qquad \alpha > 0.$$

Thus the Gabor transform $(\mathcal{G}_b^\alpha f)(w)$ localizes the Fourier transform of f around $t = b$.

$$\int_{-\infty}^{\infty} g_\alpha(t-b) db = \int_{-\infty}^{\infty} g_\alpha(x) dx = 1,$$

so that

$$\int_{-\infty}^{\infty} (\mathcal{G}_b^\alpha f)(w) db = \hat{f}(w), \qquad w \in \mathbb{R}.$$

That is, the set $\{\mathcal{G}_b^\alpha f : b \in \mathbb{R}\}$ of Gabor transform of f decomposes the Fourier transform $\hat{f}$ of f exactly, to give its local spectral information.

The width of the window function g_α is given by

$$2\Delta_{g_\alpha} = \frac{2}{\|g_\alpha\|_2} \Big\{ \int_{-\infty}^{\infty} x^2 {g_\alpha}^2(x) dx \Big\}^{1/2},$$

with centre '0'.

Theorem 1.4.3 For each $\alpha > 0$,

$$\Delta_{g_\alpha} = \sqrt{\alpha}.$$

That is, the width of the window function g_α is $2\sqrt{\alpha}$.

1.5 Wavelet transform versus Fourier transform

(i) Similarities between Fourier and wavelet transforms:

The fast Fourier transform (FFT) and the discrete wavelet transform (DWT) are both linear operations that generate a data structure that contains $\log_2 n$ segments of various lengths, usually filling and transforming it into a different data vector of length 2^n.

The mathematical properties of the matrices involved in the transforms are similar as well. The inverse transform matrix for both the FFT and the DWT is the transpose of the original. As a result, both transforms can be viewed as a rotation in function space to a different domain. For the FFT, this new domain contains basis functions that are sines and cosines. For the wavelet transform, this new domain contains more complicated basis function called wavelets, mother wavelets, or analyzing wavelets.

Both transforms have another similarity. The basis function are localized in frequency, making mathematical tools such as power spectra (how much power is contained in a frequency interval) and scalograms useful at picking out frequencies and calculating power distributions.

(ii) Dissimilarities between Fourier and wavelet transforms:

The most interesting dissimilarity between these two kinds of transform is that individual wavelet functions are localized in space whereas Fourier sine and cosine functions are not. This localization feature, along with wavelets localization of frequency, makes many functions and operators using wavelets "sparse" when transformed into the wavelet domain. This sparseness, is turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series.

One way to see the time frequency resolution differences between the Fourier transform and the wavelet transform is to look at the basis function coverage of the time-frequency plane [71]. In a windowed Fourier transform, the window is simply a square wave. The square wave window truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the WFT, the resolution of the analysis is the same at all locations in the time-frequency plane.

An advantage of wavelet transforms is that the windows vary. In order to isolate signal discontinuities, one would like to have some very short basis functions. At the same time, in order to obtain detailed frequency analysis, one would like to have some very long basis functions. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones. This happy medium is exactly what we get with wavelet transforms.

The wavelet transforms do not have a single set of basis functions like the Fourier transform, which utilizes just the sine and cosine functions. Instead, wavelet transforms have an infinite set of possible basis functions. Thus wavelet analysis provides immediate access to information that can be obscured by other time-frequency methods such as Fourier analysis.

Definition 1.5.1 A function $\psi(x) \in L^2(\mathbb{R})$ is called a wavelet or mother wavelet if the following condition is satisfied:

$$C_\psi := \int_{-\infty}^{\infty} \frac{|\hat{\psi}(w)|^2}{|w|}\, dw < \infty, \tag{1.1}$$

where, $\hat{\psi}$ is the Fourier transform of ψ.

Remark 1.5.2

(i) Condition (1.1) implies that

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{1.2}$$

Very often a function $\psi \in L^2(\mathbb{R})$ satisfying (1.2) is called a wavelet. However, in order to get analogue of inversion formula for Fourier transform we require stronger condition (1.1).

(ii) For $\psi \in L^2(\mathbb{R})$ satisfying $t\psi \in L^1(\mathbb{R})$, i.e. $\int_{-\infty}^{\infty} t\psi(t)dt < \infty$, condition (1.1) is equivalent to

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \text{ resp. } \hat{\psi}(0) = 0.$$

Lemma 1.5.3 Let ϕ be a nonzero n-times $(n \geq 1)$ differentiable function such that $\phi^{(n)} \in L^2(\mathbb{R})$. Then

$$\psi = \phi^{(n)}(x)$$

is a wavelet.

Definition 1.5.4 Relative to every basic wavelet ψ, the integral wavelet transform (IWT) on $L^2(\mathbb{R})$ is defined by

$$(W_\psi f)(b,a) := |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}\, dt, \qquad f \in L^2(\mathbb{R}). \tag{1.3}$$

Remark 1.5.5 If we consider $\psi_{b,a}(t)$ as a family of functions given by

$$\psi_{b,a}(t) \;=\; |a|^{-\frac{1}{2}}\psi\left(\frac{t-b}{a}\right), \quad a > 0, \;\; b \in \mathbb{R}$$

where ψ is a fixed function, often called mother wavelet. Then (1.3) can be written as

$$(W_\psi f)(b,a) \;=\; \langle f, \psi_{b,a}\rangle.$$

Theorem 1.5.6 Let ψ be a wavelet and ϕ a bounded integrable function, then the convolution function $\psi * \phi$ is a wavelet.

Theorem 1.5.7 (Parseval's Formula for Wavelet Transforms) Let $\psi \in L^2(\mathbb{R})$ be a wavelet. Then, for any $f, g \in L^2(\mathbb{R})$, the following formula hold:

$$\langle f, g\rangle \;=\; \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (W_\psi f)(b,a)\overline{(W_\psi g)(b,a)}\; \frac{dbda}{a^2}.$$

Theorem 1.5.8 (Inversion Formula) Let $\psi \in L^2(\mathbb{R})$ be a wavelet. Then, for any $f \in L^2(\mathbb{R})$, the following formula hold:

$$f(t) \;=\; \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (W_\psi f)(b,a)|a|^{-\frac{1}{2}}\psi\left(\frac{t-b}{a}\right)\; \frac{dbda}{a^2}.$$

Definition 1.5.9[8] A wavelet $\psi \in L^2(\mathbb{R})$ is said to have n vanishing moments if

$$\int_{-\infty}^{\infty} t^k \psi(t)\, dt \;=\; 0, \qquad \text{for } \; 0 \le k < n.$$

Definition 1.5.10 A function $\psi \in L^2(\mathbb{R})$ is called a dyadic wavelet if there exist two positive constants A and B with $0 < A \le B < \infty$ such that

$$A \;\le\; \sum_{j=-\infty}^{\infty} |\hat{\psi}(2^{-j}w)|^2 \;\le\; B \qquad \text{a.e.}$$

Definition 1.5.11

(i) Given $a > 0$, the dilation operator D_a, defined on function $f \in L^1(\mathbb{R})$, is given by

$$D_a f(x) = a^{1/2} f(ax).$$

(ii) Given $b \in \mathbb{R}$ the translation operator T_b, defined on function $f \in L^1(\mathbb{R})$ or $L^2(\mathbb{R})$, is given by

$$T_b f(x) = f(x-b).$$

(iii) Given $c \in \mathbb{R}$, the modulation operator E_c, defined on function $f \in L^1(\mathbb{R})$ or $L^2(\mathbb{R})$, is given by

$$E_c f(x) = e^{2\pi i c x} f(x).$$

Remark 1.5.12

(i) Translation means shifting a function by b.

(ii) For $a > 1, D_a f(x)$ is narrowed down version of $f(x)$, that is, D_a compresses $f(x)$. If $0 < a < 1$, then $D_a f(x)$ is spread out version of $f(x)$, that is, D_a stretches $f(x)$.

(iii) The modulation operator E_c rotates $f(x)$.

1.6 Examples of wavelets

Example 1.6.1 (Haar Wavelet) The function defined below is called the Haar function

$$\psi(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ -1, & \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

ψ has compact support, i.e. ψ is zero outside [0,1). Since $\psi \in L^2(\mathbb{R})$ and also satisfied the condition (1.1), therefore the Haar function is a wavelet.

Example 1.6.2 (Mexican Hat Wavelet) The function ψ defined as

$$\psi(x) = (1 - x^2)e^{-x^2/2}$$

is known as the Mexican hat function. It is a wavelet by Lemma 1.5.3. Mexican hat wavelet is the second derivative of Gaussian function $-e^{-x^2/2}$.

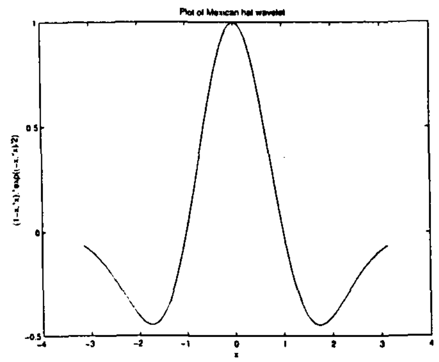

Figure 1.1: Mexican hat wavelet

Example 1.6.3 (Morlet or Gabor Wavelet) A function defined as

$$\psi(x) = (e^{i\alpha t} - e^{-\alpha^2/2})e^{-t^2/2}$$

is called Morlet or Gabor function. Since $\psi \in L^2(\mathbb{R}), x\psi(x) \in L^1(\mathbb{R})$, by Remark 1.5.2(ii), ψ is a wavelet and is known as Morlet or Gabor wavelet.

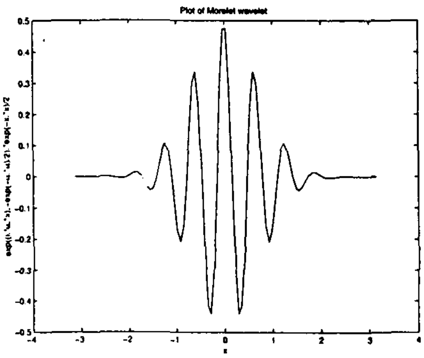

Figure 1.2: Morlet wavelet

Example 1.6.4 If we take the Haar wavelet and convolute it with the following function

$$\phi(t) = \begin{cases} 0, & t < 0 \\ 1, & 0 \le t < 1 \\ 0, & t \ge 1, \end{cases}$$

we obtain a simple wavelet.

Example 1.6.8 The convolution of the Haar wavelet with $\phi(t) = e^{-t^2}$ generates a smooth wavelet.

1.7 Construction of wavelets from a multiresolution analysis

The concept of multiresolution analysis (MRA), formulated by Y. Meyer [58] and S. Mallat [57] is crucial to the theory of wavelets.

Definition 1.7.1 A multiresolution analysis (MRA) is a sequence of closed subspaces $\{V_j; j \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ satisfying the following properties:

(i) $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots$

(ii) $\text{clos}_{L_2}(\bigcup_{j\in\mathbb{Z}} V_j) = L_2(\mathbb{R})$,

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$,

(iv) $f(x) \in V_j$ if and only if $f(2^{-j}x) \in V_0$,

(v) $f(x) \in V_0$ if and only if $f(x-m) \in V_0$ for all $m \in \mathbb{Z}$

(vi) there exists a function $\phi \in V_0$ called the scaling function such that the system $\{\phi(t-m)\}_{m\in\mathbb{Z}}$ orthonormal basis in V_0.

Remark 1.7.2

(a) Condition (i) to (iii) mean that every function in $L^2(\mathbb{R})$ can be approximated by elements of the subspaces V_j, and as j approaches ∞, the precision of approximation increases.

(b) Conditions (iv) and (v) express the invariance of the system of subspaces $\{V_j\}$ with respect to the translation and dilation operators.

(c) Condition (v) follows from (vi).

(d) Condition (vi) can be rephrased for each $j \in \mathbb{Z}$ that the system $\{2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_j.

(e) For a given MRA $\{V_j\}$ in $L^2(\mathbb{R})$ with scaling function ϕ, a wavelet is obtained in the following manner. Let the subspace W_j of $L^2(\mathbb{R})$ be defined by the condition

$$V_j + W_j = V_{j+1}, \quad V_j \perp W_j \quad \text{for all } j$$

$$V_{j+1} = J_j(V_0 \oplus W_0) = J_j(V_0) \oplus J_j(W_0) = V_j \oplus J_j(W_0),$$

where J_j (for an integer j, J_j is defined as $J_j(f(x)) = f(2^j x)$ for all $f \in L^2(\mathbb{R})$) is an isometry, $J_j(V_1) = V_{j+1}$.

$$V_m = \bigoplus_{j \geq m+1} W_j.$$

This gives

$$W_j = J_j(W_0) \quad \text{for all } j \in \mathbb{Z}.$$

From condition (i) to (iii), we obtain an orthogonal decomposition

$$L^2(\mathbb{R}) = \oplus \sum_{j \in \mathbb{Z}} W_j = W_1 \oplus W_2 \oplus W_3 \oplus \ldots.$$

$$= \bigoplus_{j \in \mathbb{Z}} W_j.$$

Let $\psi \in W_0$ be such that $\{\psi(t-m)\}_{m \in \mathbb{Z}}$ orthonormal basis in W_0, then $\{\psi_{j,k} = 2^{j/2}\psi(2^j.-m) : m \in \mathbb{Z}\}$ is an orthonormal basis of W_j. This function is a wavelet. Let $\phi(x) = \sum_{n \in \mathbb{Z}} c_n \phi(2x-n)$, where c_n is an appropriate constant, and then $\psi(x) = \sum_{n \in \mathbb{Z}} (-1) c_{n+1} \phi(2x+n)$.

(f) It may be noted that the convention of increasing subspaces $\{V_j\}$ is not universal. Very often decreasing sequences of subspaces $\{V_j\}$ are used in the definition. However one gets similar results.

Since $V_0 \subset V_1$, any function in V_0 can be expanded in terms of the basis functions of V_1. In particular, $\phi(x) = \phi_{0,0} \in V_0$ and hence

$$\phi(x) = \sum_{k=-\infty}^{\infty} a_k \phi_{1,k}(x) = \sqrt{2} \sum_{k=-\infty}^{\infty} a_k \phi(2x-k),$$

where

$$a_k = \int_{-\infty}^{\infty} \phi(x)\phi_{1,k}(x)dx. \tag{1.4}$$

For compactly supported scaling functions only finitely many a_k's will be nonzero and we have [23]

$$\phi(x) = \sqrt{2} \sum_{k=0}^{D-1} a_k \phi(2x-k). \tag{1.5}$$

Equation (1.5) is fundamental for the wavelet theory and it is known as the dilation equation. D is an even positive integer called the wavelet genus and the numbers $a_0, a_1, ..., a_{D-1}$, are called filter coefficients. The scaling function is uniquely characterized (up to a constant) by these coefficients.

In analogy to (1.5), we can write a relation for the basic wavelet ψ. Since $\phi \in W_0$ and $W_0 \subset V_1$, we expand ψ as

$$\psi(x) = \sqrt{2}\sum_{k=0}^{D-1} b_k\phi(2x-k). \tag{1.6}$$

This equation is called the wavelet equation, where the filter coefficients are given by

$$b_k = \int_{-\infty}^{\infty} \psi(x)\phi_{1,k}(x)dx. \tag{1.7}$$

Although the filter coefficients a_k and b_k are formally defined by (1.4) and (1.7) respectively, they are not normally computed that way because we do not know ϕ and ψ explicitly. However, they can be found indirectly from properties of ϕ and ψ. Note that b_k can be expressed in terms of a_k as follows:

Theorem 1.7.3 $b_k = (-1)^k a_{D-1-k}, \quad k = 0, 1, ..., D-1.$

Since $\int_{-\infty}^{\infty} \phi(x)dx = 1$. Integrating both sides of (1.5) then yields

$$\sum_{k=0}^{D-1} a_k = \sqrt{2}.$$

Taking the Fourier transform on both sides of the equation (1.5), we obtain

$$\hat{\phi}(\xi) = \sqrt{2}\sum_{k=0}^{D-1} a_k \int_{-\infty}^{\infty} \phi(2x-k)e^{-i\xi x}dx. \tag{1.8}$$

With the change of variable $2x - k = y$, we get

$$\begin{aligned}
&= \sqrt{2}\sum_{k=0}^{D-1} a_k \int_{-\infty}^{\infty} \phi(y)e^{-i\xi(y+k)/2}dy/2 \\
&= \frac{1}{\sqrt{2}}\sum_{k=0}^{D-1} a_k e^{-ik\xi/2} \int_{-\infty}^{\infty} \phi(y)e^{-i(\xi/2)y}\, dy \\
&= m_0\left(\frac{\xi}{2}\right)\hat{\phi}\left(\frac{\xi}{2}\right),
\end{aligned}$$

where

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{k=0}^{D-1} a_k e^{-ik\xi}. \tag{1.9}$$

Lemma 1.7.4 If ψ has P vanishing moments then

$$m_0(0) = 1$$

$$\frac{d^p}{d\xi^p} m_0(\xi)|_{\xi=\pi} = 0, \quad p = 0, 1, 2....., P-1.$$

For proof we refer to [61].

Corollary 1.7.5

$$m_0(n\pi) = \begin{cases} 1, & n \text{ even} \\ 0, & n \text{ odd.} \end{cases}$$

Equation (1.8) can be repeated for $\hat{\phi}(\xi/2)$ to yield

$$\hat{\phi}(\xi) = m_0\left(\frac{\xi}{2}\right) m_0\left(\frac{\xi}{4}\right) \hat{\phi}\left(\frac{\xi}{4}\right).$$

After N steps, we have

$$\hat{\phi}(\xi) = \Pi_{j=1}^{N} m_0\left(\frac{\xi}{2^j}\right) \hat{\phi}\left(\frac{\xi}{2^N}\right).$$

It follows from (1.9) that $|m_0(\xi)| \leq 1$ so that product converges for $N \to \infty$. Thus we obtain an expression

$$\hat{\phi}(\xi) = \Pi_{j=1}^{\infty} m_0\left(\frac{\xi}{2^j}\right) \hat{\phi}(0).$$

Using $\hat{\phi}(0) = \int_{-\infty}^{\infty} \phi(x)dx = 1$, we now arrive at the following product formula

$$\hat{\phi}(\xi) = \Pi_{j=1}^{\infty} m_0\left(\frac{\xi}{2^j}\right), \quad \xi \in \mathbb{R}.$$

Lemma 1.7.6

$$\hat{\phi}(2\pi n) = \delta_{0,n}, \quad n \in \mathbb{Z}.$$

Theorem 1.7.7

$$\sum_{n=-\infty}^{\infty} \phi_{j,0}(x+n) = 2^{-j/2}, \quad j \leq 0, \quad x \in \mathbb{R}.$$

For proof we refer to [61].

We can obtain an analogous for $\hat{\psi}$. Using (1.6).

$$\begin{aligned}
\hat{\psi}(\xi) &= \sqrt{2}\sum_{k=0}^{D-1} b_k \int_{-\infty}^{\infty} \phi(2x-k)e^{-i\xi x}dx \\
&= \frac{1}{\sqrt{2}}\sum_{k=0}^{D-1} b_k e^{-ik\xi/2} \int_{-\infty}^{\infty} \phi(y)e^{-i(\xi/2)y}\, dy \\
&= m_1\left(\frac{\xi}{2}\right)\hat{\phi}\left(\frac{\xi}{2}\right),
\end{aligned}$$

where

$$m_1(\xi) = \frac{1}{\sqrt{2}}\sum_{k=0}^{D-1} b_k e^{-ik\xi}. \tag{1.10}$$

Using Theorem 1.7.3 we can express $m_1(\xi)$ in terms of $m_0(\xi)$, as

$$\begin{aligned}
m_1(\xi) &= \frac{1}{\sqrt{2}}\sum_{k=0}^{D-1}(-1)^k a_{D-1-k}e^{-ik\xi} \\
&= \frac{1}{\sqrt{2}}\sum_{k=0}^{D-1} a_{D-1-k}e^{-ik(\xi+\pi)} \\
&= \frac{1}{\sqrt{2}}\sum_{l=0}^{D-1} a_l e^{-i(D-1-l)(\xi+\pi)} \\
&= e^{-i(D-1)(\xi+\pi)}\frac{1}{\sqrt{2}}\sum_{l=0}^{D-1} a_l e^{il(\xi+\pi)} \\
&= e^{-i(D-1)(\xi+\pi)}\overline{m_0(\xi/2+\pi)}\ \hat{\phi}(\xi/2).
\end{aligned}$$

The following theorem provide a relationship between scaling function, MRA and wavelets.

Theorem 1.7.8 Let $\phi \in L^2(\mathbb{R})$ satisfy

(i) $\{\phi(t-m)\}$ is a Riesz sequence of $L^2(\mathbb{R})$,

(ii) $\phi(x/2) = \sum_{k\in\mathbb{Z}} a_k\phi(x-k)$ converges on $L^2(\mathbb{R})$, and

(iii) $\hat{\phi}(\xi)$ is continuous at 0 and $\hat{\phi}(0) \neq 0$, where $\hat{\phi}$ denotes the Fourier transform of ϕ. Then the spaces $V_j = \text{span}\{\phi(2^j x-k)\}_{k\in\mathbb{Z}}$ with $j \in \mathbb{Z}$ form an MRA.

Theorem 1.7.9 Let $\{V_j\}$ be an MRA with a scaling function $\phi \in V_0$. The function $\psi \in W_0 = V_1 \ominus V_0$ $(W_0 \oplus V_0 = V_1)$ is a wavelet if and only if

$$\hat{\psi}(2\xi) = e^{i\xi} v(2\xi) \overline{m_0(\xi + \pi)} \hat{\phi}(\xi)$$

for some 2π-periodic function $v(\xi)$ such that $|v(\xi)| = 1$ almost everywhere, where $m_0(\xi) = 2^{-1/2} \sum_{k \in \mathbb{Z}} a_k e^{-ik\xi}$ is the associated low pass filter.

1.8 Wavelet packets

We have the following sequence of functions due to Wickerhauser [74]. For $n = 0, 1, 2, 3, \cdots$

$$w_{2n}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} a_k w_n(2t - k) \tag{1.11}$$

$$w_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} b_k w_n(2t - k) \tag{1.12}$$

where, the functions $w_0(t)$ and $w_1(t)$ can be identified with the functions ϕ and ψ respectively, as defined in [25], and $a = \{a_k\}$ is the filter which satisfies the following properties,

$$\sum_{n \in \mathbb{Z}} a_{n-2k} \cdot a_{n-2l} = \delta_{k,l}$$

$$\sum_{n \in \mathbb{Z}} a_n = \sqrt{2} \quad \text{and} \quad b_k = (-1)^k a_{1-k}.$$

We denote the summing and differencing operators H and G on $l^2(\mathbb{Z})$ by

$$Hf(i) = \sum_{k \in \mathbb{Z}} a_{k-2i} \cdot f(k) \ ; \ Gf(i) = \sum_{k \in \mathbb{Z}} b_{k-2i} \cdot f(k). \tag{1.13}$$

The adjoint operators H^* and G^* are defined by

$$H^* f(k) = \sum_{i \in \mathbb{Z}} a_{k-2i} \cdot f(i) \ ; \ G^* f(k) = \sum_{i \in \mathbb{Z}} b_{k-2i} \cdot f(i) \tag{1.14}$$

and

$$H^*H + G^*G = I. \tag{1.15}$$

For $n = 0$ in (1.11) and (1.12), we get,

$$w_0(t) = w_0(2t) + w_0(2t - 1)$$

$$w_1(t) = w_0(2t) - w_0(2t - 1) .$$

If we increase n, we get the following

$$w_2(t) = w_1(2t) + w_1(2t-1)$$
$$w_3(t) = w_1(2t) - w_1(2t-1)$$
$$w_4(t) = w_1(4t) + w_1(4t-1) + w_1(4t-2) + w_1(4t-3) .$$

We observe that w_n's have a "fixed scale" but different frequencies. They are Walsh functions in $[0,1[$. The function $w_n(t-k)$, for integers k, n with $n \geq 0$, form an orthonormal basis of $L^2(\mathbb{R})$ [74].

We define a space Ω_n which is the linear span of integer translates of w_n's as

$$\Omega_n := \{f \mid f = \sum_{k\in\mathbb{Z}} c_k w_n(t-k)\}, \tag{1.16}$$

where $\{c_k\} \in l^2(\mathbb{Z})$.

Using identity (1.15), we have

$$w_n(t-k) = \frac{1}{\sqrt{2}} \sum_i a_{k-2i}\, w_{2n}\left(\frac{t}{2} - i\right) + \frac{1}{\sqrt{2}} \sum_i b_{k-2i}\, w_{2n+1}\left(\frac{t}{2} - i\right) .$$

From (1.15) and (1.16),

$$f(t) = \frac{1}{\sqrt{2}} \left\{ \sum_i Hf(i)\, w_{2n}\left(\frac{t}{2} - i\right) + \sum_i Gf(i)\, w_{2n+1}\left(\frac{t}{2} - i\right) \right\}$$

or, $\sqrt{2} f(t) = p + q$ for $p \in \Omega_{2n}$ and $q \in \Omega_{2n+1}$.

We may define

$$\delta f(t) = \sqrt{2} f(2t) .$$

Therefore,

$$\delta\Omega_n = \Omega_{2n} \oplus \Omega_{2n+1}$$

or in more general form

$$\delta^k \Omega_n = \Omega_{2^k n} \oplus \cdots \oplus \Omega_{2^k(n+1)-1}; \quad k \geq 0.$$

In the construction of a wavelet from an MRA, essentially the space V_1 was split into two orthogonal components V_0 and W_0. Note that V_1 is the closure of the linear span of the functions $\{2^{1/2}\phi(2. - k) : k \in \mathbb{Z}\}$, whereas V_0 and W_0 are respectively the closure of the span of $\{\phi(. - k) : k \in \mathbb{Z}\}$ and $\{\psi(. - k) : k \in \mathbb{Z}\}$. Since $\phi(2. - k) = \phi(2(. - \frac{k}{2}))$, we see that the above procedure splits the half integer translates of a function into integer translates of two functions.

In fact, in the wavelet packets the splitting is not only confined to V_1 alone: we can choose to split W_j, which is the span of $\{\psi(2^j. - k) : k \in \mathbb{Z}\} = \{\psi(2^j(. - \frac{k}{2^j})) : k \in \mathbb{Z}\}$, to get two functions whose $2^{-(j-1)}k$ translates will span the same space W_j. Repeating the splitting procedure j times, we get 2^j functions whose integer translates alone span the space W_j. If we apply this to each W_j, then the resulting basis of $L^2(\mathbb{R})$, which will consist of integer translates of a countable number of functions (instead of all dilations and translations of the wavelet ψ), will give us a better frequency localization. This basis is called "wavelet packet basis".

Theorem 1.8.1 For every partition P of the nonnegative integers into the sets of the form $I_{kn} = \{2^k n, \cdots, 2^k(n+1) - 1\}$, the collection of functions $W^n_{k,j} = 2^{k/2} w_n(2^k t - j), I_{kn} \in P, j \in \mathbb{Z}$ is an orthonormal basis of $L^2(\mathbb{R})$.

Definition 1.8.2 Wavelet packet basis of $L^2(\mathbb{R})$ is an orthonormal basis selected from among the functions

$$\{2^{k/2} w_n(2^k t - j), \ j \in \mathbb{Z}\}.$$

1.9 Numerical calculation of wavelet packet coefficients

Let $f(t)$ be a function in $L^2(\mathbb{R})$ and let $\{c_p : p \in \mathbb{Z}\}$ be the coefficients of $f(t)$ in $\delta^L \Omega_0$. Here, $c_p = \int 2^{L/2} w_0(2^L t - p) f(t) dt$ and the L^2 function given by the projection will be denoted by

$$P_L f(t) = \sum_{p \in \mathbb{Z}} c_p \ 2^{L/2} w_0(2^L t - p).$$

From $\{c_p\}$ we may calculate the coefficients of $f(t)$ in any space $\delta^k \Omega_n$, for $0 \leq k \leq L$, and $0 \leq n < 2^{L-k}$, by applying the functions G and H to the sequences $\{c_p\}$. Thus we have

$$c_p^{fs} = \int_{-\infty}^{\infty} 2^{s/2} w_f(2^s t - p) f(t) dt$$

$p \in \mathbb{Z}, 0 \leq s \leq L, \ 0 \leq f < 2^{L-s}$. The coefficients of $f(t)$ in the subspace $\delta^k \Omega_n$ form a sequence $\{c_p^{nk} : p \in \mathbb{Z}\}$.

As for example, if we take $L = 3$, then the boxes of coefficients in the rectangle correspond to the decomposition of $\delta^3\Omega_0$ into the subspaces $\delta^k\Omega_n$, for $0 \leq k \leq 3$, and $0 \leq n < 2^{3-k}$. The top box corresponds to $\delta^3\Omega_0$, the bottom boxes correspond to Ω_n, for $0 \leq n < 2^3$, and box n on level k (counting the bottom as level 0) corresponds to subspace $\delta^k\Omega_n$ (see the following table).

<table>
<tr><td colspan="8" align="center">$\delta^3\Omega_0$</td></tr>
<tr><td colspan="4" align="center">$\delta^2\Omega_0$</td><td colspan="4" align="center">$\delta^2\Omega_1$</td></tr>
<tr><td colspan="2" align="center">$\delta\Omega_0$</td><td colspan="2" align="center">$\delta\Omega_1$</td><td colspan="2" align="center">$\delta\Omega_2$</td><td colspan="2" align="center">$\delta\Omega_3$</td></tr>
<tr><td>Ω_0</td><td>Ω_1</td><td>Ω_2</td><td>Ω_3</td><td>Ω_4</td><td>Ω_5</td><td>Ω_6</td><td>Ω_7</td></tr>
</table>

We have many choices to represent $\delta^3\Omega_0$ as direct sum of orthonormal basis subsets. The Wavelet basis $\Omega_0 \oplus \Omega_1 \oplus \delta\Omega_1 \oplus \delta^2\Omega_1$, the subband basis $\delta\Omega_0 \oplus \delta\Omega_1 \oplus \delta\Omega_2 \oplus \delta\Omega_3$, and the Walsh basis $\Omega_0 \oplus \Omega_1 \oplus \cdots \oplus \Omega_7$ are obtained. Thus, we have constructed wavelet packets which provides a family of orthonormal bases for $L^2(\mathbb{R})$. The optimal representation of the data within the library of wavelet packets is obtained by using the so called "best basis algorithm" [74].

1.10 Frames, wavelet frames and their properties

The history of frames is a nice example of the development of mathematics. Frames were introduced already in 1952 by Duffin and Schaeffer [32] in their fundamental paper, they used frames as a tool in the study of nonharmonic Fourier series, i.e. sequences of the type $\{e^{i\lambda_n x}\}_{n\in\mathbb{Z}}$, where $\{\lambda_n\}_{n\in\mathbb{Z}}$ is a family of real or complex numbers. Apparently, the importance of the concept was not realized by the mathematical community; at least it took almost 30 years before the next treatment appeared in print. In 1980, Young [80] wrote his book, which contains the basic facts about frames. Then, in 1985, as the wavelet era began, Daubechies, Grossmann and Meyer [26] observed that frames can be used to find series expansions of functions in $L^2(\mathbb{R})$ which are very similar to the expansions using orthonormal bases. This was probably the time when many mathematicians started to see the potential of the topic; this point became more clear via Daubechies [23, 24] and the combined survey/research paper by Heil and Walnut [45].

Given a function $\psi \in L^2(\mathbb{R})$ and parameters $a > 1, b > 0$, the associated wavelet system is the collection of functions $\{a^{j/2}\psi(a^j x - kb)\}_{j,k\in\mathbb{Z}}$. A frame of this type is called a wavelet frame. The definition shows that all the functions in the wavelet frame are generated by certain scaling and translations of the single function ψ. A slight generalization is to consider frames generated by scaling and translating of a finite collection of functions $\psi_1, \psi_2, ..., \psi_n$; a frame $\{a^{j/2}\psi_l(a^j x - kb)\}_{j,k\in\mathbb{Z},\ l=1,2,.....,n}$

is called a wavelet packet frame.

Definition 1.10.1 A sequence of vectors $\{x_n\}$ in a Hilbert space $\mathcal{H}$ is called a basis (Schauder basis) of $\mathcal{H}$ if to each $x \in \mathcal{H}$, there corresponds a unique sequence of scalars $\{a_n\}_{n=1}^{\infty}$ such that

$$x = \sum_{n=1}^{\infty} a_n x_n.$$

where the convergence is defined by the norm.

Definition 1.10.2 A basis $\{x_n\}_{n=1}^{\infty}$ of $\mathcal{H}$ is called orthogonal if $\langle x_n, x_m \rangle = 0$ for $n \neq m$.
An orthogonal basis is called orthonormal if $\langle x_n, x_n \rangle = 1$ for all n.

Definition 1.10.3 An orthogonal basis $\{x_n\}_{n=1}^{\infty}$ is complete in the sense that if $\langle x, x_n \rangle = 0$, for all n, then $x = 0$ [27].

Definition 1.10.4 If $\{x_n\}$ is a basis in a separable Hilbert space $\mathcal{H}$, then

(i) $\{x_n\}$ is called a bounded basis if there exist two nonnegative numbers A and B such that

$$A \leq \|x_n\| \leq B, \qquad \text{for all n.}$$

(ii) $\{x_n\}$ is called an unconditional basis if

$$\sum a_n\ x_n \in \mathcal{H} \text{ implies that} \sum |a_n|\ x_n \in \mathcal{H}.$$

(iii) $\{x_n\}$ is called a Riesz basis if there exist a topological isomorphism $T : \mathcal{H} \to \mathcal{H}$ and an orthonormal basis $\{y_n\}$ of $\mathcal{H}$ such that $Tx_n = y_n$ for every n.

Remark 1.10.5 In a Hilbert space, all bounded unconditional basis are equivalent to an orthonormal basis. In other words, if $\{x_n\}$ is a bounded unconditional basis, then there exists an orthonormal basis $\{e_n\}$ and topological isomorphism $T : \mathcal{H} \to \mathcal{H}$ such that $Te_n = x_n$ for every n.

Definition 1.10.6 A sequence $\{x_n\}$ in a separable Hilbert space $\mathcal{H}$ (not necessarily a basis of $\mathcal{H}$) is called a frame if there exist two numbers A and B with $0 < A \leq B < \infty$ such that

$$A\|x\|_2^2 \leq \sum_{n=1}^{\infty} |\langle x, x_n \rangle|^2 \leq B\|x\|_2^2, \quad \forall x \in \mathcal{H}.$$

The numbers A and B are called *frame bounds.* If $A = B$, the frame is called *tight.* The frame is called *exact* if it ceases to be a frame whenever any single element is deleted from the frame.

Definition 1.10.7 To each frame $\{x_n\}$ there corresponds an operator T, called the frame operator from $\mathcal{H}$ into itself defined by

$$Tx = \sum_n \langle x, x_n \rangle \, x_n, \quad \forall x \in \mathcal{H}.$$

Definition 1.10.8 Let $\mathcal{H}$ be a Hilbert space and $T : \mathcal{H} \to \mathcal{H}$ be a bounded linear operator on $\mathcal{H}$ into itself. The adjoint operator T^* of T is defined by

$$\langle Tx, y \rangle = \langle x, T^* y \rangle, \quad \forall \; x, y \in \mathcal{H}.$$

Remark 1.10.9

(i) T^* always exists.

(ii) T^* is bounded, linear and unique.

(iii) The adjoint operator of T^* is denoted by T^{**}.

Theorem 1.10.10 Let T be a bounded linear operator on a Hilbert space $\mathcal{H}$ into itself. Then, its adjoint operator T^* has the following properties:

(i) $I^* = I$, where I is the identity operator.

(ii) $(T + S)^* = T^* + S^*$

(iii) $(\alpha T)^* = \overline{\alpha} T^*$

(iv) $(TS)^* = S^* T^*$

(v) $T^{**} = T$

(vi) $\|T^*\| = \|T\|$

(vii) $\|T^* T\| = \|T\|^2$

(viii) If T is invertible, so is T^* and $(T^*)^{-1} = (T^{-1})^*$.

Definition 1.10.11 Let T be a bounded linear operator on a Hilbert space $\mathcal{H}$ into itself, then,

(i) T is called self adjoint or Hermitian if $T = T^*$.

(ii) A self adjoint operator T is called a positive operator if $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$. It is called strictly positive if $\langle Tx, x \rangle = 0$ only if $x = 0$.

(iii) T is called normal if $TT^* = T^* T$.

(iv) T is called unitary if $TT^* = T^*T = I$, where I is the identity operator.

Examples 1.10.12 If $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis of $\mathcal{H}$, then

(i) $\{e_1, e_1, e_2, e_2, e_3, e_3, ...\}$ is a tight frame with frame bounds $A = B = 2$, but it is not exact.

(ii) $\{\sqrt{2}e_1, e_2, e_3, ...\}$ is an exact frame but not tight since the frame bounds are easily seen as $A = 1$ and $B = 2$.

(iii) $\{e_1, \frac{e_2}{\sqrt{2}}, \frac{e_2}{\sqrt{2}}, \frac{e_3}{\sqrt{3}}, \frac{e_3}{\sqrt{3}}, \frac{e_3}{\sqrt{3}}, ...\}$ is a tight frame with the frame bound $A = B = 1$ but not an orthonormal basis.

(iv) $\{e_1, \frac{e_2}{2}, \frac{e_3}{3}, ...\}$ is a complete orthogonal sequence but is not a frame.

Theorem 1.10.13 If a sequence $\{x_n\}$ is a tight frame in $\mathcal{H}$ with the frame bound $A = 1$, and if $\|x_n\| = 1$ for all n, then $\{x_n\}$ is an orthonormal basis of $\mathcal{H}$.

Theorem 1.10.14 Suppose a sequence $\{x_n\}$ is a separable Hilbert space $\mathcal{H}$. Then, the following are equivalent:

(a) The frame operator $Tx = \sum_n \langle x, x_n\rangle\, x_n$ is a bounded linear operator on it with $AI \le T \le BI$, where I is the identity operator on $\mathcal{H}$.

(b) $\{x_n\}_{n=-\infty}^{\infty}$ is a frame with frame bounds A and B.

Theorem 1.10.15 Suppose $\{x_n\}_{n=1}^{\infty}$ is frame on a separable Hilbert space $\mathcal{H}$ with frame bounds A and B, and T is the corresponding frame operator. Then,

(a) T is invertible and $B^{-1}I \le T^{-1} \le A^{-1}I$. Furthermore, T^{-1} is a positive operator and hence it is self-adjoint.

(b) $\{T^{-1}x_n\}$ is a frame with frame bounds B^{-1} and A^{-1} with $A^{-1} \ge B^{-1} > 0$, and it is called the dual frame of $\{x_n\}$.

(c) Every $x \in \mathcal{H}$ can be expressed in the form

$$x = \sum_n \langle x, T^{-1}x_n\rangle x_n = \langle x, x_n\rangle T^{-1}x_n.$$

Theorem 1.10.16 Suppose $\{x_n\}_{n=-\infty}^{\infty}$ is a frame on a separable Hilbert space $\mathcal{H}$ with frame bounds A and B. If there exists a sequence of scalars $\{c_n\}$ such that $x = \sum_n c_n x_n$, then

$$\sum_n |c_n|^2 = \sum_n |a_n|^2 + \sum_n |a_n - c_n|^2,$$

where $a_n = \langle x, T^{-1}x_n \rangle$ so that $x = \sum_n a_n x_n$.

Theorem 1.10.17 A necessary and sufficient condition for a sequence $\{x_n\}$ on a Hilbert space $\mathcal{H}$ to be an exact frame is that the sequence $\{x_n\}$ be bounded unconditional basis of $\mathcal{H}$.

Chapter 2

Generalizations of Wavelet Packet Frame

2.1 Introduction

The essential problem in signal analysis is to find a numerically stable algorithm for reconstruction of a signal from its atomic decomposition [21]. This leads to the notion of frames [24, 45] which is a main ingredient in the analysis and synthesis of signals.

Introduced by Duffin and Schaeffer [32] in the context of non-harmonic Fourier series, the theory of frames has been developed for Gabor and Wavelets by many authors, see especially the papers by Daubechies [24], Heil and Walnut [45], Christensen [14], Sun and Zhou [66] and Shang and Zhou [64].

In this chapter, we have obtained the frame bounds for wavelet packet frames for $L^2(\mathbb{R})$ which are more general than that of wavelet frames. The frame bounds are also obtained by taking scaling function in $H^s(\mathbb{R})$

2.2 Results

Throughout this chapter we have taken the convention $\psi_{l;j,k} = 2^{j/2}\psi_l(2^j x - k), j, k \in \mathbb{Z}, l = 1, 2, ..., k$. The definition and properties of wavelet packets are already given in chapter 1.

Theorem 2.2.1 If $\psi_{l;j,k} = 2^{j/2}\psi_l(2^j x - k),\ j, k \in \mathbb{Z}, l = 1, 2, ..., k$, constitute a wavelet packet frame for $L^2(\mathbb{R})$ with frame bounds A, B, then

$$A\ ln2 \leq \int_0^\infty \xi^{-1}|\hat{\psi}_l(\xi)|^2 d\xi \leq\ B\ ln2.$$

Proof. We use the Ad^+ to denote the set of all functions whose Fourier transform is continuous and compactly supported in the positive half real line, and vanishes in some neighborhood of zero. We use P^+ to denote the right half-plane, i.e., $P^+ = \{(a,b) \in \mathbb{R}^2 : a > 0, b \in \mathbb{R}\}$. Let $w(s) = e^{-\lambda^2 \pi s^2}$, where λ is a positive parameter. We define the weight function,

$$c(a,b) = \begin{cases} w(\frac{|b|}{a}), & 1 \le a \le 2 \\ 0, & \text{otherwise}, \end{cases}$$

and define

$$C = \int_0^\infty \frac{da}{a^2} \int_{-\infty}^\infty db \langle ., h^{a,b} \rangle h^{a,b} w\left(\frac{|b|}{a}\right),$$

where, $h^{a,b} = |a|^{-1/2}\, h\left(\frac{.-b}{a}\right)$.
Since $\{\psi_{l;j,k} = 2^{j/2}\psi_l(2^j x - k)\}_{j,k \in \mathbb{Z},\ l=1,2,\ldots,k}$, constitutes a wavelet packet frame for $L^2(\mathbb{R})$ with frame bounds A, B, we have

$$A\, \|h\|^2 \le \sum_l \sum_{j,k} |\langle \psi_{l;j,k}, h^{a,b} \rangle|^2 \le B\, \|h\|^2, \text{ for } (a,b) \in P^+. \tag{2.1}$$

Multiplying equation (2.1) by weight function c(a,b) and integrating over P^+ , we have

$$A\, TrC \le \sum_l \sum_{j,k} \langle C\psi_{l;j,k}, \psi_{l;j,k} \rangle \le B\, TrC \tag{2.2}$$

where,

$$\begin{aligned} TrC &= \textstyle\int_{P^+} \frac{dadb}{a^2} c(a,b) \|h\|^2 \\ &= \textstyle\int_1^2 \frac{da}{a^2} \int_{-\infty}^\infty db\ w\left(\frac{|b|}{a}\right) \|h\|^2 \\ &= \textstyle\int_1^2 \frac{da}{a} \int_{-\infty}^\infty ds\ w(s) \|h\|^2 \\ &= ln2\ 2\left(\textstyle\int_0^\infty ds\ w(s) \|h\|^2\right). \end{aligned}$$

For weight function w we have chosen, $\int_0^\infty dtw(t) = \frac{1}{2}$.
Hence,

$$TrC = ln2\ \|h\|^2. \tag{2.3}$$

The middle term of equation (2.2), becomes

$$\textstyle\sum_l \sum_{j,k} \langle C\psi_{l;j,k}, \psi_{l;j,k} \rangle = \sum_l \sum_{j,k} \int_1^2 \frac{da}{a^2} \int_{-\infty}^\infty db\ w\left(\frac{|b|}{a}\right) |\langle \psi_{l;j,k}, h^{a,b} \rangle|^2. \tag{2.4}$$

Now,

$$\begin{aligned}
\langle \psi_{l;j,k}, h^{a,b} \rangle &= \int \psi_{l;j,k}(x)\, \overline{h^{a,b}(x)} dx \\
&= \int 2^{j/2} \psi_l(2^j x - k) a^{-1/2}\, \overline{h\left(\tfrac{x-b}{a}\right)} dx \\
&= 2^{j/2} a^{-1/2} \int \psi_l(y)\, \overline{h\left(\tfrac{y+k-b2^j}{2^j a}\right)} 2^{-j} dy \\
&= 2^{-j/2} a^{-1/2} \int \psi_l(y)\, \overline{h\left(\tfrac{y-(b2^j-k)}{2^j a}\right)} dy \\
&= \langle \psi_l, h^{2^j a, b2^j - k} \rangle.
\end{aligned}$$

Now equation(2.4) becomes,

$$\sum_l \sum_{j,k} \langle C\psi_{l;j,k}, \psi_{l;j,k} \rangle = \sum_l \sum_{j,k} \int_1^2 \tfrac{da}{a^2} \int_{-\infty}^{\infty} db w\left(\tfrac{|b|}{a}\right) |\langle \psi_l, h^{2^j a, b2^j - k} \rangle|^2.$$

By changing variables $2^j a = a'$, $2^j b = b'$, we have

$$\begin{aligned}
\sum_l \sum_{j,k} \langle C\psi_{l;j,k}, \psi_{l;j,k} \rangle &= \sum_l \sum_{j,k} \int_{2^j}^{2^{j+1}} \tfrac{da'}{a'^2} \int_{-\infty}^{\infty} db' w\left(\tfrac{|b'|}{a'}\right) |\langle \psi_l, h^{a', b'-k} \rangle|^2 \\
&= \sum_l \sum_{j\in\mathbb{Z}} \int_{2^j}^{2^{j+1}} \tfrac{da}{a^2} \int_{-\infty}^{\infty} db \sum_{k\in\mathbb{Z}} w\left(\tfrac{b+k}{a}\right) |\langle \psi_l, h^{a,b} \rangle|^2 \\
&= \sum_l \sum_{j\in\mathbb{Z}} \int_0^{\infty} \frac{da}{a^2} \int_{-\infty}^{\infty} db |\langle \psi_l, h^{a,b} \rangle|^2 \sum_{k\in\mathbb{Z}} w\left(\frac{b+k}{a}\right)
\end{aligned} \tag{2.5}$$

The function $w(s) = e^{-\lambda^2 \pi s^2}$ has only one local maximum and is monotonically decreasing as $|s|$ increases.

The Lemma 2.2 of Daubechies [24], shows that for such function w and $\alpha\beta \in \mathbb{R}$, $\beta > 0$,

$$\int_{-\infty}^{\infty} dt w(t) - \beta w_{max} \leq \beta \sum_{k\in\mathbb{Z}} w(\alpha + k\beta) \leq \int_{-\infty}^{\infty} dt w(t) + \beta w_{max},$$

or, for particular w,

$$\sum_{k\in\mathbb{Z}} w\left(\tfrac{|b+k|}{a}\right) = a + \rho(a,b)$$

with $\rho(a,b) \leq w(0) = \lambda$.
Thus we have,

$$\begin{aligned}
\sum_l \sum_{j,k} \langle C\psi_{l;j,k}, \psi_{l;j,k} \rangle &= \int_0^{\infty} \tfrac{da}{a^2} \int_{-\infty}^{\infty} db |\langle \psi_l, h^{a,b} \rangle|^2 (a + \rho(a,b)) \\
&= \int_0^{\infty} \tfrac{da}{a} \int_{-\infty}^{\infty} db |\langle \psi_l, h^{a,b} \rangle|^2 + R
\end{aligned} \tag{2.6}$$

where,

$$\begin{aligned}|R| &= \textstyle\int_0^\infty \frac{da}{a^2}\int_{-\infty}^\infty db|\langle\psi_l, h^{a,b}\rangle|^2\rho(a,b) \\ &\leq \lambda C_h\|\psi_l\|^2. \end{aligned} \tag{2.7}$$

$$\begin{aligned}\textstyle\int_0^\infty \frac{da}{a}\int_{-\infty}^\infty db|\langle\psi_l, h^{a,b}\rangle|^2 &= \textstyle\int_0^\infty \frac{da}{a}\int_{-\infty}^\infty db|\frac{1}{2\pi}\langle\hat{\psi}_l, \hat{h}^{a,b}\rangle|^2 \\ &= \textstyle\int_0^\infty \frac{da}{a}\int_{-\infty}^\infty db\frac{1}{4\pi^2}\left|\left[\int_0^\infty \hat{\psi}_l(\xi)a^{1/2}\hat{h}(a\xi)e^{ib\xi}d\xi\right]\right|^2 \\ &= \textstyle\frac{1}{2\pi}\int_0^\infty da'|\hat{h}(a')|^2\int_0^\infty d\xi\xi^{-1}|\hat{\psi}_l(\xi)|^2 \\ &= \textstyle\frac{1}{2\pi}\|\hat{h}\|_2^2\int_0^\infty d\xi\xi^{-1}|\hat{\psi}_l(\xi)|^2 \\ &= \textstyle\|h\|_2^2\int_0^\infty d\xi\xi^{-1}|\hat{\psi}_l(\xi)|^2. \end{aligned} \tag{2.8}$$

Thus equation (2.6) becomes,

$$\sum_l\sum_{j,k}\langle C\psi_{l;j,k}, \psi_{l;j,k}\rangle = \|h\|_2^2\int_0^\infty d\xi\xi^{-1}|\hat{\psi}_l(\xi)|^2 + R. \tag{2.9}$$

Substituting the values of equations (2.3),(2.9) and R in equation (2.2), we get

$$A\|h\|_2^2\, ln2 \leq \|h\|_2^2\int_0^\infty d\xi\xi^{-1}|\hat{\psi}_l(\xi)|^2 + \lambda C_h\|\psi_l\| \leq B\|h\|_2^2\, ln2.$$

Taking λ tends to zero, we get the conclusion immediately.

Theorem 2.2.2 Let $\psi_l \in L^2(\mathbb{R})$ be admissible for each l. If the system $\psi_{l;j,k} = 2^{j/2}\psi_l(2^jx-k)$, $j,k \in \mathbb{Z}, l = 1,2,....,k$, constitute a wavelet packet frame for $L^2(\mathbb{R})$ with frame bounds A, B, then

$$A\,\delta \leq \sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^j\xi)|^2 \leq B\,\Delta\,.$$

where, $\delta \leq k \leq \Delta < \infty$.

Proof. Taking the weight function,

$$c_{\lambda,\epsilon}(a,b) := \begin{cases} w_\lambda(\frac{|b|}{a}), & e^{-\epsilon} \leq a \leq e^\epsilon \\ 0, & \text{otherwise}\,, \end{cases}$$

and

$$A\;TrC_{\lambda,\epsilon,h} \leq \sum_l\sum_{j,k}\langle C_{\lambda,\epsilon,h}\psi_{l;j,k}, \psi_{l;j,k}\rangle \leq B\;TrC_{\lambda,\epsilon,h} \tag{2.10}$$

and using the similar technique as in Theorem 2.2.1, we have

$$TrC_{\lambda,\epsilon,h} \;=\; 2\|h\|^2\epsilon \tag{2.11}$$

and

$$\sum_l \sum_{j,k} \langle C_{\lambda,\epsilon,h}\psi_{l;j,k}, \psi_{l;j,k}\rangle \;=\; \sum_l \sum_{j\in\mathbb{Z}} \int_{2^j e^{-\epsilon}}^{2^j e^{\epsilon}} \frac{da}{a^2} \int_{-\infty}^{\infty} db |\langle \psi_l, h^{a,b}\rangle|^2 \sum_{k\in\mathbb{Z}} w_\lambda\left(\frac{b+k}{a}\right). \tag{2.12}$$

By Lemma 2.3 [78], we have the following estimates,

$$\frac{a}{\triangle} - \lambda \le \sum_{k\in\mathbb{Z}} w_\lambda\left(\frac{|b+k|}{a}\right) \le \frac{a}{\delta} + \frac{2\lambda\triangle}{\delta}. \tag{2.13}$$

Combining equations (2.10)-(2.13), and letting $\lambda \to 0$, we have

$$A\|h\|^2 \le \frac{1}{\delta} I_\epsilon \qquad \text{and} \qquad B\|h\|^2 \ge \frac{1}{\triangle} I_\epsilon, \tag{2.14}$$

where,

$$\begin{aligned}
I_\epsilon \;&=\; \tfrac{1}{2\epsilon}\textstyle\sum_l \sum_{j\in\mathbb{Z}} \int_{2^j e^{-\epsilon}}^{2^j e^{\epsilon}} \frac{da}{a} \int_{-\infty}^{\infty} db |\langle \psi_l, h^{a,b}\rangle|^2 \\
&=\; \tfrac{1}{4\pi^2}\tfrac{1}{2\epsilon}\textstyle\sum_l \sum_{j\in\mathbb{Z}} \int_{2^j e^{-\epsilon}}^{2^j e^{\epsilon}} \frac{da}{a} \int_{-\infty}^{\infty} db \left[\int_0^\infty |\hat{\psi}_l(\xi)|^2 a |\hat{h}(a\xi)|^2 d\xi e^{ib\xi}\right] \\
&=\; \tfrac{1}{2\pi}\textstyle\sum_l \int_0^\infty d\xi |\hat{\psi}_l(\xi)|^2 \sum_{j\in\mathbb{Z}} \frac{1}{2\epsilon} \int_{2^j e^{-\epsilon}}^{2^j e^{\epsilon}} da |\hat{h}(a\xi)|^2 \\
&=\; \tfrac{1}{2\pi}\textstyle\sum_l \int_0^\infty d\xi |\hat{\psi}_l(\xi)|^2 H_\epsilon(\xi).
\end{aligned} \tag{2.15}$$

Since $h \in Ad^+$, we can assume that supp$\hat{h} \subseteq [x_h, X_h]$ and $\hat{h}(\xi) \le M$, where x_h and X_h ($x_h < X_h$) are two positive numbers.
Here,

$$\begin{aligned}
H_\epsilon(\xi) \;&=\; \textstyle\sum_{j\in\mathbb{Z}} \frac{1}{2\epsilon} \int_{2^j e^{-\epsilon}}^{2^j e^{\epsilon}} da |\hat{h}(a\xi)|^2 \\
&=\; \textstyle\sum_{j\in\mathbb{Z}} \frac{1}{\xi 2\epsilon} \int_{\xi 2^j e^{-\epsilon}}^{\xi 2^j e^{\epsilon}} da' |\hat{h}(a')|^2 \\
&=\; \textstyle\sum_{j\in\mathbb{Z}} \frac{1}{\xi 2\epsilon} \int_{\xi 2^j e^{-\epsilon}}^{\xi 2^j e^{\epsilon}} da |\hat{h}(a)|^2 \\
&\le\; \textstyle\sum_{j\in\mathbb{Z}} \frac{M}{\xi 2\epsilon} \int_{\xi 2^j e^{-\epsilon}}^{\xi 2^j e^{\epsilon}} da \chi_{[x_h, X_h]}(a) \\
&\le\; \textstyle\sum_{\xi X_h^{-1} e^{-\epsilon} < 2^j < \xi x_h^{-1} e^{\epsilon}} \frac{M}{\xi 2\epsilon} \int_{\xi 2^j e^{-\epsilon}}^{\xi 2^j e^{\epsilon}} da \\
&\le\; \textstyle\sum_{\xi X_h^{-1} < 2^j < \xi x_h^{-1}} \frac{M}{\xi 2\epsilon} \xi 2^j (e^{-\epsilon} - e^{\epsilon}) \\
&\le\; \frac{2M}{\xi/2X_h < 2^j < 2\xi/x_h} \frac{1}{2^j}
\end{aligned}$$

$$= \quad 2M\rho_{1/2X_h,2/x_h}(\xi). \tag{2.16}$$

By Lemma 2.1 [78], the function $|\hat{\psi}_l(\xi)|^2\rho_{1/2X_h,2/x_h}(\xi)$ is integrable over $(0,\infty)$. So by the use of dominated convergence theorem, from eqn.(2.15) we can deduce that,

$$\begin{aligned}
lim_{\epsilon\to 0}I_\epsilon &= \tfrac{1}{2\pi}\textstyle\sum_l\int_0^\infty d\xi|\hat{\psi}_l(\xi)|^2H_\epsilon(\xi)\\
&= \tfrac{1}{2\pi}\textstyle\sum_l\int_0^\infty d\xi|\hat{\psi}_l(\xi)|^2 lim_{\epsilon\to 0}\tfrac{1}{2\epsilon}\sum_{j\in\mathbb{Z}}\int_{\xi 2^j e^{-\epsilon}}^{\xi 2^j e^\epsilon} da|\hat{h}(a)|^2\\
&= \tfrac{1}{2\pi}\textstyle\sum_l\int_0^\infty d\xi|\hat{\psi}_l(\xi)|^2\xi^{-1} lim_{\epsilon\to 0}\sum_{j\in\mathbb{Z}}\tfrac{1}{2\epsilon}\int_{log\xi-log2^j-\epsilon}^{log\xi-log2^j+\epsilon} du\; e^u|\hat{h}(e^u)|^2\\
&= \tfrac{1}{2\pi}\textstyle\sum_l\int_0^\infty d\xi|\hat{\psi}_l(\xi)|^2\xi^{-1}\sum_{j\in\mathbb{Z}}\tfrac{\xi}{2^j}\left|\hat{h}\left(\tfrac{\xi}{2^j}\right)\right|^2\\
&= \tfrac{1}{2\pi}\textstyle\sum_l\int_0^\infty d\xi|\hat{h}(\xi)|^2\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^j\xi)|^2\\
&= \|h\|^2\textstyle\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^j\xi)|^2.
\end{aligned} \tag{2.17}$$

Combining equations (2.14), (2.17) and Lemma 2.2 [78], the result follows.

Definition 2.2.3 A Sobolev space of order $s > 0$, denoted by $H^s(\mathbb{R})$, is a subspace of $L^2(\mathbb{R})$, given by

$$H^s(\mathbb{R}) \;=\; \left\{f\in L^2(\mathbb{R});\; |\hat{f}(\xi)|(1+|\xi|^2)^{s/2}\in L^2(\mathbb{R})\right\}.$$

Lemma 2.2.4 Suppose that the scaling function ϕ satisfies

$$\sup_{\xi\in\mathbb{R}}\sum_{k\in\mathbb{Z}}|\hat{\phi}(\xi+2k\pi)|^{2-\sigma} < +\infty, \tag{2.18}$$

for some $\sigma > 0$, and

$$\sup_{\xi\in\mathbb{R}}(1+|\xi|)^\sigma|\hat{\phi}(\xi)| < +\infty. \tag{2.19}$$

Then there exists a constant c such that, for all $f \in L^2(\mathbb{R})$,

$$\sum_l\sum_{j,k\in\mathbb{Z}}|\langle f,\psi_{l,jk}\rangle|^2 \;\leq\; c\|f\|_2^2. \tag{2.20}$$

Proof. See [19].

Theorem 2.2.5 Suppose that,

$$A \;=\; inf_{|r|\in[1,2]}\left[\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^jr)|^2-\sum_{m\neq 0}\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^jr)\hat{\psi}_l(2^jr+m)\right] > 0$$

$$B = sup_{|r|\in[1,2]}\left[\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^jr)|^2 + \sum_{m\neq0}\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^jr)\hat{\psi}_l(2^jr+m)|\right] < \infty.$$

Then $\{D_{2^j}T_k\psi_l(x)_{j,k\in\mathbb{Z},l=1,2,\ldots,k}\}$ is a wavelet packet frame for $L^2(\mathbb{R})$ with bounds A, B.

Proof. For a function $f \in L^2(\mathbb{R})$, we have

$$\begin{aligned}
\textstyle\sum_l\sum_{j,k}|\langle f, D_{2^j}T_k\psi_l\rangle|^2 &= \textstyle\sum_l\sum_{j,k\in\mathbb{Z}}|\langle \hat{f}, D_{2^{-j}}E_{-k}\hat{\psi}_l\rangle|^2 \\
&= \textstyle\sum_l\sum_{j,k\in\mathbb{Z}}|\langle \hat{f}, E_{-k2^j}D_{2^{-j}}\hat{\psi}_l\rangle|^2 \\
&= \textstyle\sum_l\sum_{j,k\in\mathbb{Z}}|\int_R \hat{f}(r)\overline{E_{-k2^j}D_{2^{-j}}\hat{\psi}_l(r)}dr|^2 \\
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}2^{-j}\sum_k|\int_R \hat{f}(r)\overline{\hat{\psi}_l(2^{-j}r)e^{2\pi ik2^{-j}r}}dr|^2 \\
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}2^{-j}2^j\int_0^{2^j}|\sum_m \hat{f}(r-2^jm)\overline{\hat{\psi}_l(2^{-j}r-m)}|^2dr
\end{aligned}$$

$$\begin{aligned}
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}\int_0^{2^j}\sum_h \hat{f}(r-2^jh)\overline{\hat{\psi}_l(2^{-j}r-h)}\sum_m\overline{\hat{f}(r-2^jm)}\hat{\psi}_l(2^{-j}r-m)dr \\
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}\sum_h\int_0^{2^j}\hat{f}(r-2^jh)\overline{\hat{\psi}_l(2^{-j}r-h)}\sum_m\overline{\hat{f}(r-2^jm)}\hat{\psi}_l(2^{-j}r-m)dr \\
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}\int_R \hat{f}(r)\overline{\hat{\psi}_l(2^{-j}r)}\sum_m\overline{\hat{f}(r-2^jm)}\hat{\psi}_l(2^{-j}r-m)dr \\
&= \textstyle\sum_l\sum_{j\in\mathbb{Z}}\sum_m\int_R \hat{f}(r)\overline{\hat{f}(r-2^jm)}\ \overline{\hat{\psi}_l(2^{-j}r)}\hat{\psi}_l(2^{-j}r-m)dr \\
&= \textstyle\int_R|\hat{f}(r)|^2\sum_l\sum_{j\in\mathbb{Z}}|\hat{\psi}_l(2^jr)|^2dr \\
&+ \textstyle\sum_{m\neq0}\sum_l\sum_{j\in\mathbb{Z}}\int_R \hat{f}(r)\overline{\hat{f}(r-2^jm)}\ \overline{\hat{\psi}_l(2^{-j}r)}\hat{\psi}_l(2^{-j}r-m)dr. \\
&= (*).
\end{aligned}$$

Applying the Cauchy-Schwarz inequality twice, we have

$$
\begin{aligned}
(*) \ \leq \ & \int_R |\hat{f}(r)|^2 \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 dr \\
& + \sum_{m\neq 0} \sum_l \sum_{j\in\mathbb{Z}} \int_R |\hat{f}(r)| \left(|\hat{\psi}_l(2^{-j}r)||\hat{\psi}_l(2^{-j}r - m)| \right)^{1/2} \\
& .|\hat{f}(r - 2^j m)| \left(|\hat{\psi}_l(2^{-j}r)||\hat{\psi}_l(2^{-j}r - m)| \right)^{1/2} dr \\
\leq \ & \int_R |\hat{f}(r)|^2 \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 dr \\
& + \sum_{m\neq 0} \sum_l \sum_{j\in\mathbb{Z}} \left(\int_R |\hat{f}(r)|^2 |\hat{\psi}_l(2^{-j}r)||\hat{\psi}_l(2^{-j}r - m)| dr \right)^{1/2} \\
& \cdot \left(\int_R |\hat{f}(r - 2^j m)|^2 |\hat{\psi}_l(2^{-j}r)||\hat{\psi}_l(2^{-j}r - m)| dr \right)^{1/2} \\
\leq \ & \int_R |\hat{f}(r)|^2 \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 dr + (a')(a'').
\end{aligned}
$$

The terms (a') and (a'') are actually identical (use the change of variable $r \to r+2^j m$ in (a'')), so by changing the summation index $j \to -j, m \to -m$, we have

$$
\begin{aligned}
(*) \ \leq \ & \int_R |\hat{f}(r)|^2 \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 dr \\
& + \sum_{m\neq 0} \sum_l \sum_{j\in\mathbb{Z}} \int_R |\hat{f}(r)|^2 |\hat{\psi}_l(2^j r)||\hat{\psi}_l(2^j r + m)| dr \\
\leq \ & \int_R |\hat{f}(r)|^2 \left(\sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 + \sum_{m\neq 0} \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)||\hat{\psi}_l(2^j r + m)| \right) dr.
\end{aligned}
$$

Thus,

$$\sum_l \sum_{j,k} |\langle f, D_{2^j} T_k \psi_l \rangle|^2 \ \leq \ B\|f\|^2.$$

A similar conclusion shows

$$(*) \ \geq \ \int_R |\hat{f}(r)|^2 \left(\sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)|^2 - \sum_{m\neq 0} \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j r)||\hat{\psi}_l(2^j r + m)| \right) dr.$$

Thus result follows.

Theorem 2.2.6 Assume that the sequence $(\alpha_n)_n$ is given by $\phi(x) = \sum_{n=n_0}^{n_1} \alpha_n \phi(2x-n)$ is finite and satisfies the condition

$$\sum_n \alpha_{n-2k}\alpha_{n-2l} \ = \ \delta_{lk}, \tag{2.21}$$

and assume that for some $\epsilon > 0$, we have

$$|\hat{\psi}(\xi)| < c(1 + |\xi|^2)^{-\epsilon-1/4}, \tag{2.22}$$

where $\hat{\psi}$ denotes the Fourier transform of the mother wavelet ψ. Finally, assume that the scaling function $\phi \in H^s(\mathbb{R})$ for some positive number s. Then there exist two positive constants $c_1(s)$ and $c_2(s)$ such that

$$\forall\, f \in L^2(\mathbb{R}), \quad c_1(s)\|f\|_2^2 \leq \sum_l \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{l,jk}\rangle|^2 \leq c_2(s)\|f\|_2^2. \tag{2.23}$$

Proof. To prove the upper bound of (2.23), it suffices to check that if $\phi \in H^s(\mathbb{R})$ for some $s > 0$, then conditions (2.18) and (2.19) of Lemma 2.2.4 are satisfied. To get (2.18) we first prove that there exists $0 < \alpha < 1$ such that

$$\int |\hat{\phi}(\xi)|^{2-2\alpha}\, d\xi < +\infty. \tag{2.24}$$

Since

$$\int |\hat{\phi}(\xi)|^{2-2\alpha}\, d\xi = \int [(1 + |\xi|)^{2s}|\hat{\phi}(\xi)|^2]^{1-\alpha}(1 + |\xi|)^{2s(\alpha-1)}\, d\xi,$$

and by using Hölder's inequality, one gets

$$\begin{aligned}\int |\hat{\phi}(\xi)|^{2-2\alpha}\, d\xi &\leq \left[\int (1 + |\xi|)^{2s}|\hat{\phi}(\xi)|^2\, d\xi\right]^{1-\alpha} \left[\int (1 + |\xi|)^{2s(\alpha-1)/\alpha}\, d\xi\right]^{\alpha} \\ &\leq c\left[\int (1 + |\xi|)^{2s(\alpha-1)/\alpha}\, d\xi\right]^{\alpha}.\end{aligned}$$

Hence, if $0 < \alpha < 1/(1 + 1/2s)$, then (2.24) holds. To get (2.18) it suffices to use the following inequalities which can be found in [19].

$$\begin{aligned}\sum_{k\in\mathbb{Z}} |\hat{\phi}(\xi + 2k\pi)|^{2-\alpha} &\leq \int \left|\frac{d}{d\xi}(|\hat{\phi}|^{2-\alpha}(\xi))\right| d\xi \\ &\leq (2-\alpha)\int \left|\frac{d\hat{\phi}}{d\xi}\right| |\hat{\phi}(\xi)|^{1-\alpha}\, d\xi \\ &\leq 2\left[\int \left|\frac{d\hat{\phi}}{d\xi}\right|^2 d\xi\right]^{1/2} \left[\int |\hat{\phi}(\xi)|^{2-2\alpha}\, d\xi\right]^{1/2}.\end{aligned} \tag{2.25}$$

Since $(\alpha_n)_n$ is a finite, it follows that the associated scaling function ϕ is compactly supported. Moreover, $\phi \in H^s(\mathbb{R})$ for some $s > 0$ implies that $\phi \in L^2(\mathbb{R})$. It becomes clear that the first factor of the last inequality is proportional to the L^2- norm of

$x\phi(x)$ which is finite, and the second factor is finite whenever $0 < \alpha < 1/(1+1/(2s))$. To prove (2.19), we consider a point ξ such that $|\xi| \in [2^{n-1}\pi, 2^n\pi]$, $n \geq 1$. Then the techniques used to get (2.25) give us

$$|\hat{\phi}(\xi)|^2 \leq c' \left[\int_{2^{n-1}\pi \leq |\xi| \leq 2^n\pi} |\hat{\phi}(\xi)|^2 \, d\xi\right]^{1/2}. \tag{2.26}$$

Since,

$$\int_{2^{n-1}\pi \leq |\xi| \leq 2^n\pi} |\hat{\phi}(\xi)|^2 \, d\xi = \int_{2^{n-1}\pi \leq |\xi| \leq 2^n\pi} \left[\left(\frac{1}{2} + |\xi|\right)^{2s} |\hat{\phi}(\xi)|^2\right] \left[\frac{1}{(1/2 + |\xi|)^{2s}}\right] d\xi,$$

therefore,

$$\begin{aligned}\int_{2^{n-1}\pi \leq |\xi| \leq 2^n\pi} |\hat{\phi}(\xi)|^2 \, d\xi &\leq c_1 \left(\frac{1}{2} + 2^{n-1}\pi\right)^{-2s} \\ &\leq c_1 \left(\frac{1}{2} + \frac{|\xi|}{2}\right)^{-2s} \\ &\leq c_2(s)(1 + |\xi|)^{-2s}.\end{aligned}$$

Consequently, there exists a constant $c_3(s)$ depending only on s such that

$$|\hat{\phi}(\xi)| \leq c_3(s)(1 + |\xi|)^{-s}, \quad \forall \xi \in R.$$

Collecting everything together, one concludes that for any arbitrary real number α satisfying

$$0 < \alpha < \min\left(s, \frac{1}{1 + 1/(2s)}\right),$$

the scaling function ϕ satisfies condition (2.18) and (2.19). Consequently, the upper bound of (2.23) is proven. To prove the lower bound of (2.23), we first mention that under condition (2.21) and (2.23), the wavelet expansion of an L^2 function f converges in the L^2-sense, that is,

$$\forall f \in L^2(\mathbb{R}), \; f(x) = \sum_l \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{l,jk} \rangle \psi_{l,jk}(x), \tag{2.27}$$

where the equality holds in the L^2-sense. For the proof of this result we refer to [25]. From (2.21) and (2.27), one concludes that, for all $f \in L^2(\mathbb{R})$,

$$\begin{aligned}\|f\|_2^2 &= \lim_{N \to +\infty} \sum_{j=-N}^{N} \sum_{k \in \mathbb{Z}} \sum_l |\langle f, \psi_{l,jk} \rangle|^2 \\ &\leq \sum_l \sum_{j,k \in \mathbb{Z}} |\langle f, \psi_{l,jk} \rangle|^2 \\ &\leq \sqrt{c_2(s)} \|f\|_2 \left[\sum_l \sum_{j,k \in \mathbb{Z}} |\langle f, \psi_{l,jk} \rangle|^2\right]^{1/2}.\end{aligned}$$

Hence,

$$\|f\|_2^2 \frac{1}{c_2(s)} \leq \sum_l \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{l,jk}\rangle|^2,$$

which proves the lower bound of (2.23) and concludes the proof of the theorem.

Theorem 2.2.7 Let $\{\psi_{l;j,k} = 2^{j/2}\psi_l(2^jx-k)\}_{j,k\in\mathbb{Z},l=1,2,....,k}$, is a wavelet packet frame of $L^2(\mathbb{R})$ with frame bounds A and B, i.e.,

$$A\,\|f\|^2 \leq \sum_l \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{l;j,k}\rangle|^2 \leq B\,\|f\|^2.$$

Then $\hat{\psi}$ satisfies

$$A \leq \sum_l \sum_{j\in\mathbb{Z}} |\hat{\psi}_l(2^j\omega)|^2 \leq B \qquad \text{a.e.},$$

for the same constants A and B.

Proof. Since $\psi_{l;j,k} = 2^{j/2}\psi_l(2^jx-k), j,k\in\mathbb{Z}, l=1,2,....,k$ and for any $f\in \mathrm{L}^2(\mathbb{R})$, we have

$$\begin{aligned}\langle f,\psi_{l;j,k}\rangle &= 2^{j/2}\int_{-\infty}^{\infty} f(x)\,\overline{\psi_l(2^jx-k)}dx\\ &= \frac{1}{2\pi}2^{j/2}\int_{-\infty}^{\infty}\hat{f}(2^j\omega)\,\overline{\hat{\psi}_l(\omega)}\,e^{ikw}d\omega.\end{aligned}$$

Now, by setting

$$T = 2\pi. \tag{2.28}$$

We have,
$\sum_l \sum_{j,k\in\mathbb{Z}} |\langle f,\psi_{l;j,k}\rangle|^2$

$$\begin{aligned}&= \sum_l\sum_{j\in\mathbb{Z}}\frac{2^j}{4\pi^2}\sum_{k\in\mathbb{Z}}\left|\int_{-\infty}^{\infty}\hat{f}(2^j\omega)\,\overline{\hat{\psi}_l(\omega)}\,e^{ikw}d\omega\right|^2\\ &= \sum_l\sum_{j\in\mathbb{Z}}\frac{2^jT^2}{4\pi^2}\sum_{k\in\mathbb{Z}}\left|\frac{1}{T}\int_0^T\left[\sum_{h\in\mathbb{Z}}\hat{f}(2^j(\omega+hT))\,\overline{\hat{\psi}_l(\omega+hT)}\right]\,e^{ik\frac{2\pi}{T}w}d\omega\right|^2\\ &= \sum_l\sum_{j\in\mathbb{Z}}2^j\int_0^T\left|\sum_{h\in\mathbb{Z}}\hat{f}(2^j(\omega+hT))\,\overline{\hat{\psi}_l(\omega+hT)}\right|^2 d\omega.\end{aligned} \tag{2.29}$$

Now, by the definition of wavelet packet frame and equation (2.29), we have

$$A \left\|\hat{f}\right\|^2 \leq \sum_l \sum_{j\in\mathbb{Z}} 2^j \int_0^T \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega + hT)) \, \overline{\hat{\psi}_l(\omega + hT)} \right|^2 d\omega \leq B \left\|\hat{f}\right\|^2. \quad (2.30)$$

If for any $M > 0, M \in Z$, and $\omega_0 \in (-\infty, \infty)$, we have

$$\sum_l \sum_{j=-M}^{M} 2^j \int_{2^{-j}\omega_0 - \frac{T}{2}}^{2^{-j}\omega_0 + \frac{T}{2}} \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega + hT)) \, \overline{\hat{\psi}_l(\omega + hT)} \right|^2 d\omega \leq B \left\|\hat{f}\right\|^2.$$

Now, consider $\hat{f} = (\frac{1}{\sqrt{2\epsilon}})\chi_{[\omega_0 - \epsilon, \omega_0 + \epsilon]}$, $\epsilon > 0$. Then for sufficiently small ϵ, the above inequality becomes

$$\sum_l \sum_{j=-M}^{M} \frac{2^j}{2\epsilon} \int_{2^{-j}(\omega_0 - \epsilon)}^{2^{-j}(\omega_0 + \epsilon)} |\hat{\psi}_l(\omega)|^2 d\omega \leq B,$$

and thus, by taking $\epsilon \to 0$ and $M \to \infty$, we have

$$\sum_l \sum_{j=-M}^{M} |\hat{\psi}_l(2^j\omega)|^2 \leq B \qquad \text{a.e.} \quad (2.31)$$

On the other hand, for any $\omega_0, \eta > 0$, a positive integer M may be chosen so that

$$\int_{2^{M+1}\omega_0/3}^{\infty} |\hat{\psi}_l(\omega)|^2 < \eta. \quad (2.32)$$

Also, for

$$0 < \epsilon < min\left\{\frac{\omega_0}{3}, \frac{T}{2}\right\},$$

the function $\hat{f} = (\frac{1}{\sqrt{2\epsilon}})\chi_{[\omega_0 - \epsilon, \omega_0 + \epsilon]}$ satisfies

$$\hat{f}(2^j(\omega + hT)) = 0$$

for all $h \in \mathbb{Z}$ with $|h| \geq (\frac{\epsilon}{2^j T}) + 1$ and all $\omega \in [2^{-j}\omega_0 - \frac{T}{2}, 2^{-j}\omega_0 + \frac{T}{2}]$. Hence, for this $\hat{f}$, we have

$$\sum_l \sum_{j=-\infty}^{-M} 2^j \int_{2^{-j}\omega_0 - \frac{T}{2}}^{2^{-j}\omega_0 + \frac{T}{2}} \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega + hT)) \, \overline{\hat{\psi}_l(\omega + hT)} \right|^2 d\omega$$

$$\le \sum_l \sum_{j=-\infty}^{-M} \frac{2^j}{2\epsilon} \int_{2^{-j}\omega_0-\frac{T}{2}}^{2^{-j}\omega_0+\frac{T}{2}} \left[\sum_{h\in\mathbb{Z}} |\hat{\psi}_l(\omega+hT)|^2 \chi_{[\omega_0-\epsilon,\omega_0+\epsilon]}(2^j(\omega+hT)) \right] (\frac{\epsilon}{2^jT}+1)$$

$$\le C \sum_l \sum_{j=-\infty}^{-M} \int_{2^{-j}(\omega_0-\epsilon)}^{2^{-j}(\omega_0+\epsilon)} \left\{ |\hat{\psi}_l(\omega)|^2 + \frac{2^j}{2\epsilon}|\hat{\psi}_l(\omega)|^2 \right\} d\omega. \tag{2.33}$$

Since $\epsilon < \frac{\omega_0}{3}$, the intervals,

$$[2^{-j}(\omega_0-\epsilon), 2^{-j}(\omega_0+\epsilon)], \qquad j \in \mathbb{Z}$$

are mutually disjoint, and hence by equation (2.32), we have

$$\sum_l \sum_{j=-\infty}^{-M} \int_{2^{-j}(\omega_0-\epsilon)}^{2^{-j}(\omega_0+\epsilon)} |\hat{\psi}_l(\omega)|^2 \le \int_{2^{M+1}\frac{\omega_0}{3}}^{\infty} \hat{\psi}_l(\omega)|^2 d\omega < \eta,$$

Now by equation (2.33), we have

$$\sum_l \sum_{j\in\mathbb{Z}} 2^j \int_0^T \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega+hT))\ \overline{\hat{\psi}_l(\omega+hT)} \right|^2 d\omega$$

$$\le \sum_l \sum_{j=-M+1}^{\infty} 2^j \int_0^T \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega+hT))\ \overline{\hat{\psi}_l(\omega+hT)} \right|^2 d\omega$$

$$+C_\eta + \frac{C}{2\epsilon} \int_{\omega_0-\epsilon}^{\omega_0+\epsilon} \sum_l \sum_{j=-\infty}^{-M} |\hat{\psi}_l(2^{-j}\omega)|^2 d\omega. \tag{2.34}$$

Therefore, by the definition of wavelet packet frame and equation (2.34), we have

$$R = \sum_l \sum_{j=-M+1}^{\infty} 2^j \int_0^T \left| \sum_{h\in\mathbb{Z}} \hat{f}(2^j(\omega+hT))\ \overline{\hat{\psi}_l(\omega+hT)} \right|^2 d\omega$$

$$\ge A - C_\eta - \frac{C}{2\epsilon} \int_{(\omega_0-\epsilon)}^{(\omega_0+\epsilon)} \sum_l \sum_{j=-\infty}^{-M} |\hat{\psi}_l(2^{-j}\omega)|^2 d\omega. \tag{2.35}$$

On the other hand , for all sufficient small $\epsilon > 0$, it is clear that

$$R = \sum_l \sum_{j=-M+1}^{\infty} 2^j \int_{2^{-j}(\omega_0-\epsilon)}^{2^{-j}(\omega_0+\epsilon)} |\hat{f}(2^j\omega)\ \overline{\hat{\psi}_l(\omega)}|^2 d\omega$$

$$= \frac{1}{2\epsilon} \int_{\omega_0-\epsilon}^{\omega_0+\epsilon} \sum_{l} \sum_{j=-M+1}^{\infty} |\hat{\psi}_l(2^{-j}\omega)|^2 d\omega,$$

where, $\hat{f} = (\frac{1}{\sqrt{2\epsilon}})\chi_{[\omega_0-\epsilon,\omega_0+\epsilon]}$. Hence , in view of the boundedness property in equation (2.31), we may take $\epsilon \to 0$ in equation (2.35) to arrive at

$$\sum_{l} \sum_{j=-M+1}^{\infty} |\hat{\psi}_l(2^{-j}\omega_0)|^2 \geq A - C_\eta - C \sum_{l} \sum_{j=-\infty}^{-M} |\hat{\psi}_l(2^{-j}\omega_0)|^2 \tag{2.36}$$

for almost all $\omega_0 > 0$. Since $\eta > 0$ is arbitrary, so from equations (2.31) and (2.36), we get

$$\sum_{l} \sum_{j \in \mathbb{Z}} |\hat{\psi}_l(2^{-j}\omega)|^2 \geq A \tag{2.37}$$

for almost all $\omega_0 > 0$. Hence , by equations (2.31) and (2.37), we get the desired result.

Chapter 3

Vector Valued Weyl-Heisenberg Wavelet Frame

3.1 Introduction

In the real world, it is often the case that different scales are needed to characterize physical properties. In his paper, EI Naschie [33, 34] while elaborating application of E-infinity contorian space time theory has rightly pointed out that "everything we see or measure is resolution dependent". E-infinity is supposed to model real space time at all resolutions including the quantum.

On the other hand, in dealing with unsteady incompressible flows, one often encounters problems whose solutions contain localized features, in which their location vary with time [76]. Arneodo et al. [6] in a very comprehensive paper emphasized the need of characterizing the scaling properties of fractal objects arising in a variety of physical situation.

To address the above, though we have a tool in the form of Fourier transform to consider the motion in the domain of frequencies but it does not perform well with the signals of time varying spectra. Even short time Fourier transform (STFT), a technique of windowing the signal with sliding windows of fixed size, does not fulfill our purpose. If one is not interested in the time of occurrence of frequency components, but only interested in frequency components, then Fourier transform could be a suitable tool. The wavelet transform on the hand, is capable of providing the time and frequency information simultaneously, hence giving the time-frequency representation of a signal.

The wavelet theory has been studied extensively by many researchers, see especially the papers by Daubechies [25], Mallat [56], Meyer [58] and Wickerhauser [74]. Most of the research papers have focused on scalar-valued wavelets or a single mother wavelet function. It is well known that there is a limitation for the time-frequency localization of a single mother wavelet [24]. Geronimo et al. [39] constructed two functions $\psi_1(x)$ and $\psi_2(x)$ whose dilations and translations form an orthonormal basis for $L^2(\mathbb{R})$. The importance of these two functions is that they are continuous, well time-localized and of certain symmetry. This illustrates the usefulness of multiwavelets, i.e., if several mother wavelets are used in an expansion, then better properties can be achieved over single wavelets. A general scheme for constructing symmetric and antisymmetric compactly supported orthonormal multiscaling functions and multiwavelets is introduced in Chui et al. [17].

Xia and Suter [79] introduced the notion of vector valued wavelets and showed that multiwavelets can be generated from the component functions in vector valued wavelets. Therefore, studying vector valued wavelets is useful in multiwavelet theory and representation of signals as well. A typical example of such vector valued signals is video images. The Moving Picture Expert Group (MPEG) can take advantage of this vector transform for image coding.

Sun and Cheng [67] investigated the construction of a class of compactly supported orthogonal vector valued wavelets. The definition and construction of orthogonal vector valued wavelet packets are given in a paper by Chen and Cheng [13].

An important aspect of wavelet transform is to obtain the good decompositions that are adapted to a given problem. The continuous decompositions are also closely related to the Coherent States (CS) approach [43]. It is well known that an elegant formulation of the theory of CS can be obtained through the group representation theory. The simplest example, namely, the wavelets and Gabor functions are obtained from the affine group '$ax + b$' and the Weyl-Heisenberg group respectively.

A tool combining the affine and the Weyl-Heisenberg groups was proposed in [69, 70]. In this case, usual representations of such a bigger group are no more square integrable, as in the wavelet case. Only the restriction of those representations to a suitable quotient space (phase-spaces) of the group restores the square integrability and thus leads to a system of CS. In this chapter, we study a new system of CS the "*vector valued Weyl-Heisenberg wavelets*", since the group considered here is the Weyl-Heisenberg with a dilation parameter.

Parallel to the continuous study of Gabor and wavelet analysis, a fundamental work on the discretization of CS was done in [23, 24]. Instead of encoding the continuous parameters of the group, one is restricted to its discrete values. The essential problem in signal analysis is to find a numerically stable algorithm for reconstruction of a signal from its atomic decomposition [21]. This leads to the notion of frames [14, 23, 32, 45] which is a main ingredient in the analysis and synthesis of signals.

The chapter is organized as follows. In Sections 3.2 and 3.3, we describe notations and the vector valued multiresolution analysis. In Section 3.4, we discuss the Weyl-Heisenberg wavelets. In the last section, we present the generic construction of frames and prove a result related to the corresponding frame bounds.

3.2 Notations and vector valued function spaces

Throughout this chapter, we use the following notations. $\mathbb{R}$ and C denote all real and complex numbers, respectively. $\mathbb{Z}$ denotes all integers.

$$L^2(\mathbb{R}, C^N) := \{\mathbf{f} = (f_1(t), f_2(t), f_3(t), ..., f_N(t))^T \ : t \in \mathbb{R}, f_k(t) \in L^2(\mathbb{R}), \ k = 1, 2, ..., N\},$$

where T means the transpose and C^N denote the N-dimensional complex Euclidean space.

$$L^2(\mathbb{R}, C^{N\times N})$$

$$:= \left\{ \mathbf{f(t)} = \begin{pmatrix} f_{11}(t) & f_{12}(t) \ .... & f_{1N}(t) \\ f_{21}(t) & f_{22}(t) \ .... & f_{2N}(t) \\ \\ f_{N1}(t) & f_{N2}(t) \ ..... & f_{NN}(t) \end{pmatrix} : t \in \mathbb{R}, f_{kl}(t) \in L^2(\mathbb{R}), k, l = 1, 2, ..., N \right\}.$$

The signal space $L^2(\mathbb{R}, C^{N\times N})$ (or $L^2(\mathbb{R}, C^N)$) is called vector valued function space. The spaces $L^2([a,b], C^N)$ and $L^2([a,b], C^{N\times N})$ are defined similarly by replacing the real line $\mathbb{R}$ with an interval [a,b]. Examples of vector valued signals are video images, where$f_{kl}(t)$ represents the pixel at the time t of the kth row and the lth column.

For $\mathbf{f} \in L^2(\mathbb{R}, C^{N\times N})$ (or $\mathbf{f} \in L^2(\mathbb{R}, C^N)$), $\|\mathbf{f}\|$ denotes the norm of $\mathbf{f}$ as

$$\|\mathbf{f}\| := \left(\sum_{k,l=1}^{N} \int_{\mathbb{R}} |f_{kl}(t)|^2 dt \right)^{1/2}$$

$$\left(\text{or} \quad \|\mathbf{f}\| := \left(\sum_{k=1}^{N} \int_{\mathbb{R}} |f_k(t)|^2 dt \right)^{1/2} \right). \tag{3.1}$$

For $\mathbf{f} \in L^2(\mathbb{R}, C^{N\times N})$, its integration $\int \mathbf{f}(t)dt$ is defined as

$$\int \mathbf{f}(t)dt := \begin{pmatrix} \int f_{11}(t)dt & \int f_{12}(t)dt \ ..... & \int f_{1N}(t)dt \\ \int f_{21}(t)dt & \int f_{22}(t)dt \ ..... & \int f_{2N}(t)dt \\ \\ \int f_{N1}(t)dt & \int f_{N2}(t)dt \ ..... & \int f_{NN}(t)dt \end{pmatrix}$$

The Fourier transform of $\mathbf{f}$ is defined here by

$$\hat{\mathbf{f}}(w) = \int_{\mathbb{R}} \mathbf{f}(t) e^{-itw} dt.$$

Then, the inverse Fourier transform is

$$\mathbf{f}(t) = \int_{\mathbb{R}} \hat{\mathbf{f}}(w) e^{itw} dt.$$

For two vector valued functions $\mathbf{f}, \mathbf{g} \in L^2(\mathbb{R}, C^{N \times N})$, $\langle f, g \rangle$ denotes the integration of the matrix product $\mathbf{f}(t)\mathbf{g}^\dagger(t)$

$$\langle f, g \rangle := \int_{\mathbb{R}} \mathbf{f}(t)\mathbf{g}^\dagger(t) dt \tag{3.2}$$

where † means the transpose and the complex conjugate. For convenience, we still call the operation $\langle , \rangle$ in (3.2) inner product although it is not the inner product in the common sense that requires to be scalar valued. The one in (3.2) is matrix valued, but still satisfies the properties for an inner product: the linearity $\langle f, a_1 g_1 + a_2 g_2 \rangle = a_1^* \langle f, g_1 \rangle + a_2^* \langle f, g_2 \rangle$ and the commutativity $\langle f, g \rangle = \langle g, f \rangle^\dagger$.

A sequence $\Phi_k(t) \in L^2(\mathbb{R}, C^{N \times N}), k \in \mathbb{Z}$ is called an orthonormal set in $L^2(\mathbb{R}, C^{N \times N})$ if

$$\langle \Phi_k, \Phi_l \rangle = \delta(k - l) I_N \tag{3.3}$$

where $\delta(k) = 1$ when $k = 0$ and $\delta(k) = 0$ when $k \neq 0$ and I_N is the $N \times N$ identity matrix. A sequence $\Phi_k(t) \in L^2(\mathbb{R}, C^{N \times N}),\ k \in \mathbb{Z}$ is called an orthonormal basis for $L^2(\mathbb{R}, C^{N \times N})$ if it satisfies (3.3), and moreover for any $\mathbf{f}(t) \in L^2(\mathbb{R}, C^{N \times N})$ there exists a sequence of $N \times N$ constant matrices F_k such that

$$\mathbf{f}(t) = \sum_{k \in \mathbb{Z}} F_k \Phi_k(t), \quad \text{for } \ t \in \mathbb{R} \tag{3.4}$$

where the multiplication $F_k \Phi_k(t)$ for each fixed t is the $N \times N$ matrix multiplication, and the convergence for the infinite summation is in the sense of the norm $\|.\|$ defined by (3.1) for the vector valued signal space. As an example of orthonormal basis for $L^2([0, 2\pi], C^{N \times N})$

$$\Phi_k(t) = \mathrm{diag}(e^{i(k_1+k)t}, e^{i(k_2+k)t}, ..., e^{i(k_N+k)t}), \quad k \in \mathbb{Z} \tag{3.5}$$

where $k_1, k_2, ..., k_N$ are N fixed integers and diag means diagonal matrix and $e^{i(k_l+k)t}$ is the lth diagonal element. It is clear that if $\{\Phi_k(t)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R}, C^{N \times N})$, then $\{U\Phi_k(t)\}_{k \in Z}$ or $\{\Phi_k(t)U\}_{k \in \mathbb{Z}}$ is also an orthonormal basis for $L^2(\mathbb{R}, C^{N \times N})$, where U is an $N \times N$ unitary matrix, i.e., $UU^\dagger = I_N$.

Let $\Phi_k(t),\ \ k \in \mathbb{Z}$, be an orthonormal basis for $L^2(\mathbb{R}, C^{N \times N})$. Then the expansion (3.4) for any $\mathbf{f} \in L^2(\mathbb{R}, C^{N \times N})$ is unique, where

$$F_k = \langle \mathbf{f}, \Phi_k \rangle, \quad k \in \mathbb{Z}. \tag{3.6}$$

We also call the expansion (3.4) the Fourier expansion of $\mathbf{f}$. The corresponding Parseval equality is

$$\langle \mathbf{f}, \mathbf{f} \rangle = \sum_{k \in \mathbb{Z}} F_k F_k^{\dagger}. \tag{3.7}$$

From (3.7) it is clear that $\langle \mathbf{f}, \mathbf{f} \rangle = 0$ if and only if $\mathbf{f} = 0$ where 0 is the zero matrix.

Although $\Phi_k(t)$ in (3.5) form an orthonormal basis, the Fourier expansion of a signal $\mathbf{f}$ is equivalent to the Fourier expansions of all independent components in $\mathbf{f}$ and the correlation between these components cannot be taken into account. It is because each $\Phi_k(t)$ is in diagonal form. Therefore, we need to seek a nondiagonal orthonormal basis for $L^2(\mathbb{R}, C^{N \times N})$.

3.3 Vector valued multiresolution analysis and vector valued wavelets

In this section, we first define vector valued multiresolution analysis for $L^2(\mathbb{R}, C^{N \times N})$. We then study its properties and construction. We also see a connection with multiwavelets.

Definition 3.3.1 A VMRA of $L^2(\mathbb{R}, C^{N \times N})$ is a nested sequence of closed subspaces $V_j,\ \ j \in \mathbb{Z}$ of $L^2(\mathbb{R}, C^{N \times N})$ such that

(1) $V_j \subset V_{j+1},\ \ j \in \mathbb{Z}$.

(2) $\cup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}, C^{N \times N})$ and $\cap_{j \in \mathbb{Z}} V_j = \{0\}$, where 0 is the zero matrix.

(3) $\mathbf{f}(t) \in V_j$ if and only if $\mathbf{f}(2t) \in V_{j+1},\ \ j \in \mathbb{Z}$.

(4) There is a $\Phi \in V_0$ such that its translations $\Phi(t) := \Phi(t-k),\ \ k \in \mathbb{Z}$, form an orthonormal basis for V_0.

The above definition for a VMRA is notationally similar to the one for conventional multiresolution analysis (MRA) [56, 57]. We call $\Phi(t)$ a vector valued scaling function (or simply scaling function) for the VMRA $\{V_j\}$. Since $\Phi(t) \in V_0 \subset V_1$, by the above definition and (3.4) there exists constant $N \times N$ matrices $H_k,\ \ k \in \mathbb{Z}$ such that

$$\Phi(t) = 2 \sum_k H_k \Phi(2t-k). \tag{3.8}$$

Let

$$H(w) = \sum_k H_k e^{ikw}. \tag{3.9}$$

Then

$$\hat{\Phi}(w) = H\left(\frac{w}{2}\right) \hat{\Phi}\left(\frac{w}{2}\right) = H\left(\frac{w}{2}\right) H\left(\frac{w}{4}\right) .....\hat{\Phi}(0). \tag{3.10}$$

In the following, without loss of generality we assume $\hat{\Phi}(0) = I_N$. Therefore

$$\hat{\Phi}(w) = H\left(\frac{w}{2}\right) H\left(\frac{w}{4}\right) ..... = \Pi_{k=1}^{\infty} H\left(\frac{w}{2^k}\right). \tag{3.11}$$

Equation (3.10) implies

$$H(0) = I_n, \quad \text{or} \quad \sum_k H_k = I_N. \tag{3.12}$$

Notice that the property (3.12) does not apply to the multiwavelets studied in [30, 31, 39].

By the orthonormality of $\Phi_k(t) = \Phi(t-k)$ (or the orthonormality of VMRA V_j), we have, in the time domain

$$\int \Phi(t)\Phi^\dagger(t+k)dt = I_N\delta(k), \quad k \in \mathbb{Z}$$

or equivalently, in the frequency domain

$$\frac{1}{(2\pi)^2} \int \hat{\Phi}(w)\hat{\Phi}^\dagger(w)e^{i2\pi k}dw = I_N\delta(k), \quad k \in \mathbb{Z}.$$

Sum them up in terms of k and interchange the order of the summation and the integration, as follows:

$$\frac{1}{2\pi} \int \hat{\Phi}(w)\hat{\Phi}^\dagger(w)\frac{1}{2\pi}\sum_k e^{i2\pi k}dw = I_N.$$

In other words

$$\frac{1}{2\pi} \int \hat{\Phi}(w)\hat{\Phi}^\dagger(w)\sum_k \delta(w+2k\pi)dw = I_N.$$

This proves that

$$\sum_k \hat{\Phi}(w+2\pi k)\hat{\Phi}^\dagger(w+2\pi k) = 2\pi I_N, \quad \forall\, w \in \mathbb{R}. \tag{3.13}$$

On the other hand, it is not hard to see the identity (3.13) implies the orthonormality of $\Phi_k(t)$ by reversing the above steps of proof. The equalities (3.10) and (3.13) also imply (similar to single wavelet) that

$$H(w)H^\dagger(w) + H(w+\pi)H^\dagger(w+\pi) = I_N, \quad \forall\, w \in \mathbb{R}. \tag{3.14}$$

For any $w \in \mathbb{R}$, let $G(w)$ satisfy

$$G(w)H^\dagger(w) + G(w+\pi)H^\dagger(w+\pi) = 0 \tag{3.15}$$

and

$$G(w)G^\dagger(w) + G(w+\pi)G^\dagger(w+\pi) = I_N. \tag{3.16}$$

Let

$$\hat{\Psi}(w) = G\left(\frac{w}{2}\right)\hat{\Phi}\left(\frac{w}{2}\right). \tag{3.17}$$

We now have the following result on the existence of a vector valued wavelet function.

Proposition 3.3.2 Let $\Psi(t)$ be the vector valued function with its Fourier transform (3.17). Then, its translations $\Psi_k(t) := \Psi(t-k),\ \ k \in \mathbb{Z}$, form an orthonormal basis for $W_0 := V_1 \ominus V_0$.

Proof. By (3.17), for $w \in \mathbb{R}$

$$\begin{aligned}
&\textstyle\sum_l \hat{\Psi}(w+2l\pi)\hat{\Psi}^\dagger(w+2l\pi) \\
&= \sum_l G\left(\frac{w}{2}+\pi l\right)\hat{\Phi}\left(\frac{w}{2}+\pi l\right)\hat{\Phi}^\dagger\left(\frac{w}{2}+\pi l\right)G^\dagger\left(\frac{w}{2}+\pi l\right) \\
&= G\left(\frac{w}{2}\right)\left(\sum_l \hat{\Phi}\left(\frac{w}{2}+2\pi l\right)\hat{\Phi}^\dagger\left(\frac{w}{2}+2\pi l\right)\right)G^\dagger\left(\frac{w}{2}\right) \\
&+ G\left(\frac{w}{2}+\pi\right)\left(\sum_l \hat{\Phi}\left(\frac{w}{2}+\pi+2\pi l\right)\hat{\Phi}^\dagger\left(\frac{w}{2}+\pi+2\pi l\right)\right)G^\dagger\left(\frac{w}{2}+\pi\right) \\
&= 2\pi\left(G\left(\frac{w}{2}\right)G^\dagger\left(\frac{w}{2}\right) + G\left(\frac{w}{2}+\pi\right)G^\dagger\left(\frac{w}{2}+\pi\right)\right) \quad \text{(by (3.13))} \\
&= 2\pi I_N \quad \text{(by (3.16))}.
\end{aligned}$$

This proves the orthonormality of $\Psi_k(t),\ \ k \in \mathbb{Z}$. We now prove its completeness.

It is clear that, for any $\mathbf{f} \in W_0$ there are constant matrices F_k such that

$$\mathbf{f}(t) = 2\sum_k F_k\Phi(2t-k).$$

Thus

$$\hat{\mathbf{f}}(w) = F\left(\frac{w}{2}\right)\hat{\Phi}\left(\frac{w}{2}\right). \tag{3.18}$$

where $F(w) = \sum_k F_k e^{ikw}$. On the other hand, $\mathbf{f} \notin V_0$ and $\mathbf{f} \in W_0$ implies

$$\int_{\mathbb{R}} \mathbf{f}(t)\Phi_k^\dagger(t)dt = 0, \quad k \in \mathbb{Z}.$$

This is equivalent to

$$\sum_l \hat{\mathbf{f}}(w+2l\pi)\hat{\Phi}^\dagger(w+2l\pi) = 0, \quad w \in \mathbb{R}$$

by using the same technique in proving the identity (3.13) from the orthonormality of $\Phi_k(t)$. From (3.10) and (3.18)

$$\sum_l F\left(\frac{w+2l\pi}{2}\right)\hat{\Phi}\left(\frac{w+2l\pi}{2}\right)\hat{\Phi}^\dagger\left(\frac{w+2l\pi}{2}\right)H^\dagger\left(\frac{w+2l\pi}{2}\right) = 0, \quad w \in \mathbb{R}.$$

Since $F(w)$ and $H(w)$ are periodic functions with period 2π, the left hand side of the above identity is

$$F\left(\frac{w}{2}\right)\left(\sum_l \hat{\Phi}\left(\frac{w}{2}+2\pi l\right)\hat{\Phi}^\dagger\left(\frac{w}{2}+2\pi l\right)\right)H^\dagger\left(\frac{w}{2}\right)$$
$$+F\left(\frac{w}{2}+\pi\right)\left(\sum_l \hat{\Phi}\left(\frac{w}{2}+\pi+2\pi l\right)\hat{\Phi}^\dagger\left(\frac{w}{2}+\pi+2\pi l\right)\right)H^\dagger\left(\frac{w}{2}+\pi\right).$$

Therefore, by the identity (3.13), we have

$$F(w)H^\dagger(w) + F(w+\pi)H^\dagger(w+\pi) = 0, \; w \in \mathbb{R}. \tag{3.19}$$

Let $F_1(w) = (F(w), F(w+\pi))^\dagger$, $H_1(w) = (H(w), H(w+\pi))^\dagger$ and $G_1(w) = (G(w), G(w+\pi))^\dagger$. Then, the identities (3.15) and (3.16) imply that, for any $w \in \mathbb{R}$, the column vectors in the $2N \times N$ matrix $H_1(w)$ and the column vectors in the $2N \times N$ matrix $G_1(w)$ are orthogonal and all these vectors form an orthonormal basis for the $2N$ dimensional complex Euclidean space C^{2N}. The identity (3.19) implies that the column vectors in the $2N \times N$ matrix $F_1(w)$ and the column vectors in $H_1(w)$ are orthogonal. Thus, there exists an $N \times N$ matrix $P(w)$ where all entries are functions of w such that

$$F(w) = P(w)G(w), \; w \in \mathbb{R}.$$

Therefore, by (3.18)

$$\hat{\mathbf{f}}(w) = P\left(\frac{w}{2}\right)G\left(\frac{w}{2}\right)\hat{\Phi}\left(\frac{w}{2}\right) = P\left(\frac{w}{2}\right)\hat{\Psi}(w).$$

By the orthonormality of $\Psi_k(t)$

$$\int_{\mathbb{R}} \hat{\mathbf{f}}(2w)\hat{\mathbf{f}}^\dagger(2w)dw = \pi\int_0^{2\pi} P(w)P^\dagger(w)dw.$$

This proves that $P(w)$ has Fourier series expansion. Let constant $N \times N$ matrices P_k be its Fourier coefficients. Then

$$\mathbf{f}(t) = \sum_k P_k\Psi(t-k).$$

This proves the completeness.

3.3.1 Connection with multiwavelets

Let $(\Phi(t))_{lk}, (\Psi(t))_{lk}$ and $(V_j)_{lk}$ be the components at the lth column and kth row of $\Phi(t), \Psi(t)$ and V_j; respectively, $l, k = 1, 2, ..., N$ and $j \in \mathbb{Z}$. Then

$$(V_j)_{lk} \subset (V_{j+1})_{lk}, \quad \text{and} \quad f(t) \in (V_j)_{lk} \Leftrightarrow f(2t) \in (V_{j+1})_{lk}$$

and $\bigcap_{j\in\mathbb{Z}}(V_j)_{lk} = \{0\}$, and $\bigcup_{j\in\mathbb{Z}} (V_j)_{lk}$ is dense in $L^2(\mathbb{R})$.
Moreover, for any $f_{lk} \in (V_0)_{lk}$, there exist constant $a_{k_1,m,l,k}$ such that

$$f_{lk}(t) = \sum_{k_1\in\mathbb{Z}} \sum_{m=1}^{N} a_{k_1,m,l,k}(\Phi(t - k_1))_{mk}, \; t \in \mathbb{R}. \tag{3.20}$$

In addition, for any $f \in L^2(\mathbb{R})$, there exist constants $a_{j,k_1,l,k}$ such that

$$f(t) = \sum_{j,k_1\in\mathbb{Z}} \sum_{l=1}^{N} a_{j,k_1,l,k}(\Psi_{jk_1}(t))_{lk}, \; t \in \mathbb{R} \tag{3.21}$$

where k is any integer with $l \le k \le N$.

Proposition 3.3.3[79] If a VQMF $H(w)$ satisfies (3.12) and (3.14), $H_k = H_k^\dagger$ for all integers k (or $H(w) = H^\dagger(w)$ for all w) and there exist a constant $C > 0$ and an integer K_0 such that for any $w \in (-2^K\pi, 2^K\pi)$ and any $K > K_0$

$$\left\|\Pi_{l=1}^{K} H\left(\frac{w}{2^l}\right)\right\| \le C \left\|\Pi_{l=1}^{\infty} H\left(\frac{w}{2^l}\right)\right\|$$

then, the solution $\Phi(t)$ in the matrix dilation equation (3.8) is a vector valued scaling function for a VMRA. Thus, the corresponding $\Psi_{jk}(t), \;\; j, k \in \mathbb{Z}'$ form an orthonormal basis for $L^2(\mathbb{R}, C^{N\times N})$.

Proposition 3.3.4[79] Let $\Phi(t)$ be a vector valued scaling function of a VMRA V_j and $\Psi(t)$ be its an associated vector valued wavelet function. Then, for any fixed $k, \;\; 1 \le k \le N$, the functions $(\Phi(t))_{lk}, \;\; l = 1, 2, ..., N$, form multiscaling functions and $(\Psi(t))_{lk}, \;\; l = 1, 2, ..., N$, form multiwavelets.

Moreover, for each pair (l, k), the spaces $(V_j)_{lk}, \;\; j \in \mathbb{Z}$, form a multiresolution analysis of multiplicity r_k where r_k is the maximum number of linearly independent functions of $(\Phi(t))_{lk}, \;\; l = 1, 2, ..., N$.

For details about multiresolution analysis of multiplicity r, we refer to ([41, 42, 46]).

A straightforward case for Proposition 3.3.3 is that $H(w)$ can be diagonalized as

$$H(w) = U^T \mathrm{diag}(\lambda_1(w), \ldots.., \lambda_N(w))U \tag{3.22}$$

where U is a constant unitary matrix and $\lambda_i(w)$ are eigenvalues of $H(w)$. In this case, if

$$\inf_{|w|<\pi/2}|\lambda_l(w)| > 0 \tag{3.23}$$

$\lambda_l(0) = 1$ and

$$|\lambda_l(w)|^2 + |\lambda_l(w+\pi)|^2 = 1 \tag{3.24}$$

where $l = 1, 2, ..., N$, then $H(w)$ generates a vector valued scaling function for a VMRA.

It is clear that if $H(w)$ has the form (3.22), then it can generate a vector valued scaling function if and only if all $\lambda_l(w)$ generates scalar valued scaling functions. Let $\phi_{(l)}(t)$ denote the scaling function generated from a lowpass quadrature mirror filter $\lambda_l(w)$, $l = 1, 2, ..., N$. Then

$$\Phi(t) = U^T \mathrm{diag}(\phi_{(1)}(t), \ldots.., \phi_{(N)}(t))U. \tag{3.25}$$

This tells us that if $\phi_{(l)}(t)$, $l = 1, 2, ..., N$, are N scalar valued scaling functions and U is a unitary matrix then $\Phi(t)$ in (3.25) is a vector valued scaling function. Let $U = (u_{kl})$. Then

$$(\Phi(t))_{kl} = \sum_m u^*_{km} u_{ml} \phi_{(m)}(t). \tag{3.26}$$

This proves the following result.

Proposition 3.3.5 Let $\Phi(t)$, $l = 1, 2, ..., N$, be N scalar valued scaling functions. Then, for any fixed k, $1 \le k \le N$, the functions $(\Phi(t))_{lk} = \sum_m u^*_{km} u_{ml} \phi_{(m)}(t)$, $l = 1, 2, ..., N$, form multiscaling functions.

Example 3.3.6 Let $\phi(t)$ be scalar valued scaling function,

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

and

$$\Phi(t) = \frac{1}{2} \begin{pmatrix} \phi(t) + \phi(-t) & \phi(t) - \phi(-t) \\ \phi(t) - \phi(-t) & \phi(t) + \phi(-t) \end{pmatrix}.$$

is a vector valued scaling function and moreover $(\Phi(t))_{11} = (\phi(t) + \phi(-t))/2$ and $(\Phi(t))_{12} = (\phi(t) - \phi(-t))/2$ form a multiscaling function. One can see that $(\Phi(t))_{11}$ is even and $(\Phi(t))_{12}$ is odd. Both of them are of certain symmerticity.

3.4 Vector valued Weyl-Heisenberg wavelets

Throughout we have taken $\Psi(t), \mathbf{f}(t) \in L^2(\mathbb{R}, C^N)$. There is slight notational difference between $\Psi(t)$ and $\mathbf{f}(t)$. Here $\Psi(t) = (\psi_1(t), \psi_2(t), \ldots.., \psi_N(t))^T : t \in \mathbb{R}, \psi_k(t) \in L^2(\mathbb{R}), k = 1, 2, \ldots., N$.
Let G denote a locally compact group, μ the left invariant Haar measure of G and U a unitary irreducible representation (UIR) of G on some Hilbert space $\mathcal{H}$. Then, we have the following definition.

Definition 3.4.1 The UIR 'U' of G is square integrable if there exists a vector $\Psi \in \mathcal{H}$ such that the following relation holds:

$$c_\Psi = \frac{1}{\|\Psi\|^2} \int_G d\mu(g) |\langle U(g)\Psi, \Psi\rangle|^2 < \infty, \; g \in G. \tag{3.27}$$

If Ψ satisfies (3.27), then it is called an admissible vector in $\mathcal{H}$.

The relation (3.27) implies that, $W_\Psi : \mathcal{H} \to L^2(G, d\mu)$, given by

$$(W_\Psi \mathbf{f})(g) = \frac{1}{\sqrt{c_\Psi}} \langle U(g)\mathbf{f}, \Psi\rangle$$

is an isometry map from $\mathcal{H}$ onto some subspace of $L^2(G, d\mu)$.

The condition (3.27) is too stringent and even in the well known cases, like the canonical group $G = \mathbb{R}^{2n} \times \mathbb{R}$, the multidimensional wavelet group $\mathbb{R}^n \times (\mathbb{R}^*_+ \times \mathrm{SO}(n))$, the Euclidean group $\mathbb{R}^{n+1} \times \mathrm{SO}(n+1)$, and the Poincaré group $\mathbb{R}^{n+1} \times \mathrm{SO}(n+1)$, the Definition 3.4.1 is not appropriate, since it leads to divergent integrals. The notion of square integrable modulo the subgroup is introduced. Let, as before, G be a locally compact group, H its closed subgroup, $X = G/H$ a homogeneous space, μ a quasi invariant on X and σ (*a Borel section*): $X = G/H \to G$. Then, we have the following definition.

Definition 3.4.2 The UIR 'U' of G in $\mathcal{H}$ is square integrable modulo the subgroup H if the following integral holds:

$$\frac{1}{\|\Psi\|^2} \int_X d\mu(x) \; |\langle U(\sigma(x))\Psi, \Psi\rangle|^2 < \infty, \; \Psi \in \mathcal{H}. \tag{3.28}$$

The section σ is called admissible.

The direct approach consists in choosing σ in such a manner that (3.28) holds. Let us apply this technique to the affine Weyl-Heisenberg group.

3.4.1 Weyl-Heisenberg wavelets

The one dimensional *affine Weyl-Heisenberg group*, consists of the set $G = \mathbb{R}^*_+ \times \mathbb{R}^2 \times \mathbb{R}$ equipped with the multiplication law,

$$
\begin{aligned}
g.g' &= (a, \underline{b}, \underline{v}; \varphi)(a', \underline{b}', \underline{v}'; \varphi') \\
&= (aa', a\underline{b}' + \underline{b}, a^{-1}\underline{v}'\underline{v}; \varphi' + \varphi + \underline{b}' \cdot a\underline{v}),
\end{aligned} \tag{3.29}
$$

where, $a \in \mathbb{R}_+^*$, $(\underline{b}, \underline{v}) \in \mathbb{R}^2$ and $\varphi \in \mathbb{R}$.

The relation (3.29) can be easily obtained as a usual matrix multiplication if we represent element $g = (a, \underline{b}, \underline{v}; \varphi) \in G$ in matrix realization by,

$$
g = (a, \underline{b}, \underline{v}; \varphi) \equiv \begin{pmatrix} 1 & a\underline{v} & \varphi \\ 0 & a & \underline{b} \\ 0 & 0 & 1 \end{pmatrix} .
$$

We easily see that the identity element is given by $e = (1, \underline{0}, \underline{0}; 0)$ and that,

$$
g^{-1} = (a^{-1}, -a^{-1}\underline{b}, -a\underline{v}; -\varphi + \underline{b} \cdot \underline{v}) .
$$

This group G is *unimodular*, i.e., the left and right Haar measures coincide,

$$
d\mu_R(g) = d\mu_L(g) = d\mu(g) = \frac{1}{a} da\, d\underline{b}\, d\underline{v}\, d\varphi .
$$

We now consider the *Stone-Von-Neumann representation* of G on $L^2(\mathbb{R}, C^N)$, defined in [51], by

$$
[U(a, \underline{b}, \underline{v}; \varphi)\mathbf{f}]\,(\underline{x}) = \mathbf{f}_{(a,\underline{b},\underline{v};\varphi)}(\underline{x}) = \frac{1}{\sqrt{a}} e^{i[\varphi + \underline{v}\cdot(\underline{x} - \underline{b})]} \mathbf{f}(a^{-1}(\underline{x} - \underline{b})) . \tag{3.30}
$$

We use the Fourier transform of a function $\mathbf{f} \in L^2(\mathbb{R})$, defined by

$$
\hat{\mathbf{f}}(\underline{k}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbf{f}(\underline{x}) e^{-i\underline{k}\cdot\underline{x}} d\underline{x} .
$$

Then, the relation (3.30) is rewritten equivalently as follows:

$$
\left[U(a, \underline{b}, \underline{v}; \varphi)\hat{\mathbf{f}}\right](\underline{k}) = \sqrt{a}\, e^{i[\varphi - \underline{k}\cdot\underline{b}]} \hat{\mathbf{f}}(a(\underline{k} - \underline{v})) . \tag{3.31}
$$

Let us consider the *one-parameter subgroup*, as defined in [70], by

$$
H = \{h \in G \mid h = (a, \underline{0}, \underline{0}; \varphi)\} .
$$

Now, a more general section from the corresponding homogeneous space $X = G/H$ has of the form:

$$
\sigma(x) = (1, \underline{b}, \underline{v}; 0)(\alpha(\underline{b}, \underline{v}), \underline{0}, \underline{0}; \beta(\underline{b}, \underline{v})) = (\alpha(\underline{b}, \underline{v}), \underline{b}, \underline{v}; \beta(\underline{b}, \underline{v})), \tag{3.32}
$$

where, $\alpha = \alpha(\underline{b}, \underline{v})$ and $\beta = \beta(\underline{b}, \underline{v})$ are continuous functions on the phase-space X with a generic point $(\underline{b}, \underline{v})$.

The restriction of the representation U to X in (3.31) is given by

$$\left[U(\sigma(\underline{b}, \underline{v})\hat{\mathbf{f}}\right](\underline{k}) = \sqrt{\alpha(\underline{b}, \underline{v})}e^{i[\beta(\underline{b},\underline{v})-\underline{k}\cdot\underline{b}]}\hat{\mathbf{f}}(\alpha(\underline{b}, \underline{v})(\underline{k} - \underline{v})) .$$

Let us associate to any $\mathbf{f} \in L^2(\mathbb{R}, C^N)$ the following family of coefficients:

$$\left(W_{\alpha\beta}\mathbf{f}\right)(\underline{b}, \underline{v}) = \langle \mathbf{f}, \mathbf{\Psi}^{\alpha\beta}_{(\underline{\mathbf{b}},\underline{\mathbf{v}})}\rangle,$$

where,

$$\Psi^{\alpha\beta}_{(\underline{b},\underline{v})}(\underline{x}) = \frac{1}{\sqrt{\alpha(\underline{b}, \underline{v})}}e^{i[\beta(\underline{b},\underline{v})+\underline{v}\cdot(\underline{x}-\underline{b})]}\Psi(\alpha^{-1}(\underline{b}, \underline{v})(\underline{x} - \underline{b})) .$$

Now, the direct approach consists in reconstructing $\mathbf{f}(\underline{\mathbf{x}})$ from the coefficients of $\left(W_{\alpha\beta}\mathbf{f}\right)(\underline{b}, \underline{v})$, as

$$\mathbf{F}(\underline{x}) = \int_X d\underline{b}\, d\underline{v}\, \Psi^{\alpha\beta}_{(\underline{b},\underline{v})}(\underline{x}) .$$

Using the Fourier transform and assuming that α is a function depending only on $\underline{v}$, i.e., $\alpha = \alpha(\underline{v})$, we get

$$\hat{\mathbf{F}}(\underline{k}) = \hat{\mathbf{f}}(\underline{k}) \int_{\mathbb{R}} d\underline{v}\, |\hat{\Psi}(\alpha(\underline{v})(\underline{k} - \underline{v}))|^2\alpha(\underline{v}) . \tag{3.33}$$

By the change of integration variable, $\overset{\circ}{\underline{k}} = \overset{\circ}{\underline{k}}(\underline{v}) = \alpha(\underline{v})(\underline{k} - \underline{v})$, the transformation (3.33) leads to the resolution of the identity ($\hat{\mathbf{F}} = \alpha(\Psi)\hat{\mathbf{f}}$) if and only if $\alpha'(\underline{v})\alpha^{-2}(\underline{v})$ is a constant function, i.e.,

$$\alpha(\underline{v}) = \frac{1}{\underline{\lambda} \cdot \underline{v} + \mu},\ \underline{v} \neq -\mu\underline{\lambda}^{-1}, \tag{3.34}$$

where, $\underline{\lambda}, \mu$ are constants of integration. The corresponding admissibility condition is given by

$$d(\Psi) = \int_{\mathbb{R}} |\hat{\Psi}(\underline{k})|^2 \frac{d\underline{k}}{|1 + \underline{\lambda} \cdot \underline{k}|} < \infty . \tag{3.35}$$

Thus, $\underline{\lambda}$ in (3.34) implies that the dilation parameter is constant, leading to the standard *Gabor analysis* or *usual windowed Fourier transform.* More explicitly, the following constitutes a family of coherent states:

$$\left\{\Psi_{(\underline{b},\underline{v})}(\underline{x}) = \mu^{\frac{1}{2}}e^{i\underline{v}\cdot(\underline{x}-\underline{b})}\Psi\left(\mu(\underline{x} - \underline{b})\right)\right\},$$

for a constant μ and $\Psi \in \mathcal{H} = L^2(\mathbb{R}, C^N)$.

The Gabor function here is a dilated copy of Ψ, that is, letting $\mu = 1$, we get the pure windowed Fourier transform. On the other hand, $\underline{\lambda} \neq 0$ implies that, from (3.34),

$$\underline{v} = (\alpha^{-1} - \mu)\frac{\underline{\lambda}}{|\underline{\lambda}|^2} \equiv (\alpha^{-1} - \mu)\underline{\lambda}^{-1}, \ d\underline{b}\, d\underline{v} = \frac{d\underline{b}\, da}{a^2},$$

which gives the *wavelet analysis*,

$$\Psi_{(\underline{b},\underline{v})}(\underline{x}) = (\mu + \underline{\lambda}\cdot\underline{v})^{\frac{1}{2}} e^{i\underline{v}\cdot(\underline{x}-\underline{b})}\Psi\left((\mu + \underline{\lambda}\cdot\underline{v})(\underline{x} - \underline{b})\right). \tag{3.36}$$

In *Fourier variables*,

$$\hat{\Psi}_{(\underline{b},\underline{v})}(\underline{k}) = (\mu + \underline{\lambda}\cdot\underline{v})^{-\frac{1}{2}} e^{i\underline{k}\cdot\underline{b}}\hat{\Psi}\left((\mu + \underline{\lambda}\cdot\underline{v})^{-1}(\underline{k} - \underline{v})\right) . \tag{3.37}$$

We get the usual *wavelet analysis* by letting $\mu = 0$ and $|\underline{\lambda}| >> 1$. That is,

$$\hat{\Psi}_{(\underline{b},\underline{v})}(\underline{k}) = a^{\frac{1}{2}} e^{i\underline{k}\cdot\underline{b}}\hat{\Psi}(a\underline{k}) .$$

3.5 Vector valued Weyl-Heisenberg wavelet frames

This section is devoted to the discussion on some aspects of non-orthogonal discrete affine vector valued Weyl-Heisenberg wavelet expansions, parallel to the usual wavelets and windowed Fourier transforms [24].

3.5.1 Atomic decomposition

Let us consider the lattice in $\mathbb{R}^2$ for some values of $\underline{\lambda}$ and μ as:

$$G_{m,n} = \{(m,n) \in \mathbb{Z}^2 \mid \underline{b} = a_0^m\, n\, \underline{b}_0, \underline{v} = \underline{\lambda}^{-1}(a_0^{-m} - \mu)\},$$

where, a_0, b_0 are real fixed numbers. Notice that, according to (3.34),

$$a_0^m = a_0^m(\underline{\lambda}, \mu) = \frac{1}{\mu + \underline{\lambda}\cdot\underline{v}} .$$

Thus,

$$\lim_{\underline{\lambda}\to 0} a_0^m(\underline{\lambda}, \mu) = \mu^{-1}, \ \lim_{\underline{\lambda}\to 0} \underline{v}(\underline{\lambda}, \mu) = m\,\underline{v}_0 < \infty,$$

$$\lim_{\mu\to 0} \underline{v}(\underline{\lambda}, \mu) = \underline{\lambda}^{-1}\, a_0^{-m} \text{ and } \lim_{\mu\to 0} a_0^m(\underline{\lambda}, \mu) = (\underline{\lambda}\cdot\underline{v})^{-1} \equiv a_o^m .$$

Recall that $G_{m,n}$ has no more group structure. According to (3.37), we have

$$\hat{\Psi}_{m,n}(\underline{k}) = a_0^{\frac{m}{2}}\, e^{i\underline{k}\cdot n\,\underline{b}_0\, a_0^m}\,\hat{\Psi}\left(a_0^m(\underline{k} + \mu\,\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right). \tag{3.38}$$

Here, as in the continuous case, we get

$$\lim_{\underline{\lambda}\to 0, \mu\to 1} \hat{\Psi}_{m,n}(\underline{k}) = e^{i\underline{k}\cdot n\,\underline{b}_0}\hat{\Psi}(\underline{k} - m\underline{v}_0)$$

and this is the discretely labelled family of Gabor transform. Similar is readily obtained in the wavelet limit with the condition $|\lambda| \gg 1$, i.e.,

$$\lim_{\mu \to 0, |\lambda| \gg 1} \hat{\Psi}_{m,n}(\underline{k}) = a_0^{\frac{m}{2}} e^{i\underline{k} \cdot n \underline{b}_0 a_0^m} \hat{\Psi}(a_0^m \underline{k}) \ .$$

Now, the fundamental problem is to see if affine Weyl-Heisenberg coefficients $\langle \mathbf{f}, \Psi_{m,n} \rangle$ characterize a function $\mathbf{f}$ or are sufficient to reconstruct $\mathbf{f}$. If any function $\mathbf{f}$ can be written as such superposition, functions $\Psi_{m,n}$ are called "*atoms*" and the corresponding expansions $\langle \mathbf{f}, \Psi_{m,n} \rangle$ the "*atomic decomposition*". These questions are closely related to the existence of frames, the notion of which is briefly recalled for further investigations.

3.5.2 Frame bounds

The integrability condition (3.28) may be written in the following fashion by

$$\int_X d\mu(x) |\Psi_x\rangle\langle\Psi_x| = \mathbb{A}, \tag{3.39}$$

where, $\mathbb{A}$ is bounded linear operator on $\mathcal{H}$ with bounded inverse,

$$\Psi_x = U(\sigma(x))\Psi, \quad \Psi \in \mathcal{H},$$

σ is an admissible section.

Definition 3.5.1 The set of vectors $\Psi_x \in \mathcal{H}, x \in X = G/H$ is a frame if:

(i) for every $x \in X$, $\{\Psi_x\}$ is a linearly independent set.

(ii) there exists $\mathbb{A}$ in (3.39) such that the integral converges weakly.

This is, in fact, equivalent to say

$$m(\mathbb{A})\|\mathbf{f}\|^2 \leq \int_X |\langle \Psi_x \mid \mathbf{f} \rangle|^2 d\mu(x) \leq M(\mathbb{A})\|\mathbf{f}\|^2, \forall \mathbf{f} \in \mathcal{H}, \tag{3.40}$$

where, $m(\mathbb{A})$ and $M(\mathbb{A})$ are the infimum and the supremum respectively, of the spectrum of $\mathbb{A}$ and are called the frame bounds.

We define the dual frame or basis as $\tilde{\Psi}_x = \mathbb{A}^{-1}\Psi_x$ and the frame condition (3.40) is satisfied with frame bounds $M(\mathbb{A})^{-1}$ and $m(\mathbb{A})^{-1}$. Thus the function $\mathbf{f}$ may be reconstructed using the dual basis $\tilde{\Psi}_x$ as

$$\mathbf{f}(y) = \int_X d\mu(x) \langle \mathbf{f} \mid \Psi_x \rangle \tilde{\Psi}_x(y) \ .$$

This notion of continuous frame is taken from the theory of non-orthogonal expansions or atomic decompositions. Indeed, if X is discrete and μ a counting measure, we recover the classical definition of a frame [23] from (3.40),

$$A\|\mathbf{f}\|^2 \leq \sum_{x_0} |\langle \Psi_{x_0} \mid \mathbf{f}\rangle|^2 \leq B\|\mathbf{f}\|^2, x_0 \in X_0 \subset \mathbb{Z}^d, \tag{3.41}$$

where, d is the range of the lattice X_0. In our case, $d = 2$, and the sequence

$$\langle \Psi_{m,n} \mid \mathbf{f}\rangle_{m,n\in\mathbb{Z}} \in l^2(\mathbb{Z}^2) .$$

If the sequence, $\{\Psi_{m,n}\}_{m,n\in\mathbb{Z}}$ be a frame in $\mathcal{H}$, then the operator, $\mathbf{F} : \mathcal{H} \to l^2(\mathbb{Z}^2)$, given by

$$(\mathbf{F}\,\mathbf{f})_{m,n} = \langle \mathbf{f}, \Psi_{m,n}\rangle$$

is called the *frame operator.* The operator $\mathbf{F}$ is bounded according to the second inequality in (3.41). Its dual $\mathbf{F}^*$ is easily computed by

$$\langle \mathbf{F}^*\, c, \mathbf{f}\rangle = \langle c, \mathbf{F}\mathbf{f}\rangle = \big\langle \sum_{m,n} c_{m,n}\Psi_{m,n}, \mathbf{f}\big\rangle,$$

where, $c = \{c_{m,n}\}_{m,n\in\mathbb{Z}} \in l^2(\mathbb{Z}^2)$.

We assume $\Psi \in \mathcal{H} = L^2(\mathbb{R}, C^N)$ to be admissible in the sense of (3.35). It is necessary to estimate the frame bounds corresponding to the decomposition by (3.38). The relation (3.41) may be written in the following operational form,

$$A\,\mathbf{1} \leq \mathbf{F}^*\,\mathbf{F} \leq B\,\mathbf{1} .$$

The admissibility condition of Ψ is not enough to guarantee the frames of wavelets. The following proposition imposes some conditions on $\Psi, a_0, \underline{b}_0$ under which we do indeed obtain a frame, and we estimate the corresponding frame bounds.

Proposition 3.5.2 If Ψ, a_0 are such that

$$\inf_{|\underline{k}|\leq a_0} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|^2 > 0,$$

$$\sup_{|\underline{k}|\leq a_0} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|^2 < \infty,$$

and if

$$\gamma_{\underline{\lambda},\mu}(\underline{\xi}) = \sup_{\underline{k}\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|\,|\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} - \underline{\xi})\right)|$$

decays as fast as $\left(1+|\underline{\lambda}^{-1}-\underline{\xi}|\right)^{-(1+\epsilon)}$, with $\epsilon>0$, then there exists a critical $b_0^\Omega>0$ such that the $\Psi_{m,n}$ constitutes a frame for $b_0<b_0^\Omega$ and the frame bounds are:

$$\begin{aligned} A(\underline{\lambda},\mu) &= \frac{2\pi}{\underline{b}_0}\left\{\inf_{|\underline{k}|\le a_0}\sum_{m\in\mathbb{Z}}|\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)|^2\right.\\ &\left.\quad-\sum_{l\in\mathbb{Z},l\neq 0}\left[\gamma_{\underline{\lambda},\mu}\left(\frac{2\pi}{\underline{b}_0}l\right)\gamma_{\underline{\lambda},\mu}\left(-\frac{2\pi}{\underline{b}_0}l\right)\right]^{\frac{1}{2}}\right\}\end{aligned}$$

and

$$\begin{aligned} B(\underline{\lambda},\mu) &= \frac{2\pi}{\underline{b}_0}\left\{\sup_{|\underline{k}|\le a_0}\sum_{m\in\mathbb{Z}}|\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)|^2\right.\\ &\left.\quad+\sum_{l\in\mathbb{Z},l\neq 0}\left[\gamma_{\underline{\lambda},\mu}\left(\frac{2\pi}{\underline{b}_0}l\right)\gamma_{\underline{\lambda},\mu}\left(-\frac{2\pi}{\underline{b}_0}l\right)\right]^{\frac{1}{2}}\right\}.\end{aligned}$$

Proof. Consider the expression,

$$\begin{aligned}\sum_{m,n}|\langle\Psi_{m,n}\,|\,\mathbf{f}\rangle|^2 &= \sum_{m,n}\int_{\mathbb{R}\times\mathbb{R}} d\underline{k}\,d\underline{k}'\;a_0^m\,e^{[i(\underline{k}-\underline{k}')\,n\underline{b}_0 a_0^m]}\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)\\ &\quad\times\overline{\hat{\Psi}\left(a_0^m(\underline{k}'+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)\hat{\mathbf{f}}(\underline{k})}\hat{\mathbf{f}}(\underline{k}')\,. \end{aligned}\tag{3.42}$$

Let us consider the Poisson formula,

$$\sum_{n\in\mathbb{Z}}e^{2\pi ikqn}=\frac{2\pi}{q}\sum_{l\in\mathbb{Z}}\delta(k-\frac{l}{q})\,.\tag{3.43}$$

With the use of (3.43), the relation (3.42) becomes,

$$\begin{aligned}&\textstyle\sum_{m,n}|\langle\Psi_{m,n}\,|\,\mathbf{f}\rangle|^2\\ &\quad=\frac{2\pi}{\underline{b}_0}\sum_{m,l}\int_{\mathbb{R}}d\underline{k}\;\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)\\ &\qquad\times\overline{\hat{\Psi}\left(a_0^m(\underline{k}+\frac{2\pi\,l}{\underline{b}_0a_0^m}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)\hat{\mathbf{f}}(\underline{k})}\hat{\mathbf{f}}\left(\underline{k}+\frac{2\pi\,l}{\underline{b}_0a_0^m}\right)\\ &\quad=\frac{2\pi}{\underline{b}_0}\int_{\mathbb{R}}d\underline{k}\;|\hat{\mathbf{f}}(\underline{k})|^2\sum_m|\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1})-\underline{\lambda}^{-1}\right)|^2\\ &\qquad+\mathrm{Rest}(\mathbf{f})\end{aligned}\tag{3.44}$$

where,

$$\text{Rest}(\mathbf{f}) = \frac{2\pi}{\underline{b}_0} \sum_{m,l\in\mathbb{Z},l\neq 0} \int_{\mathbb{R}} d\underline{k}\, \hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right) \overline{\times \hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) + (\frac{2\pi\, l}{\underline{b}_0} - \underline{\lambda}^{-1})\right)\hat{\mathbf{f}}(\underline{k})}\, \hat{\mathbf{f}}\left(\underline{k} + \frac{2\pi\, l}{\underline{b}_0 a_0^m}\right). \quad (3.45)$$

By the Cauchy-Schwarz inequality and the variable change, $\tilde{\underline{k}} = \underline{k} - \frac{2\pi\, l}{\underline{b}_0 a_0^m}$ in (3.45),

$$\begin{aligned} |Rest(\mathbf{f})| \;\leq\; & \frac{2\pi}{\underline{b}_0} \sum_{m,l\in\mathbb{Z},l\neq 0} \left[\int_{\mathbb{R}} d\underline{k}\, |\hat{\mathbf{f}}(\underline{k})|^2 |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)| \right. \\ & \left. \times |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} - \frac{2\pi\, l}{\underline{b}_0})\right)|\right]^{\frac{1}{2}} \\ & \times \left[\int_{\mathbb{R}} d\tilde{\underline{k}}\, |\hat{\mathbf{f}}(\tilde{\underline{k}})|^2 |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)| \right. \\ & \left. \times |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} + \frac{2\pi\, l}{\underline{b}_0})\right)|\right]^{\frac{1}{2}}. \end{aligned}$$

Using the same Cauchy-Schwarz inequality for the summation over m, we get

$$\begin{aligned} |Rest(\mathbf{f})| \;\leq\; & \frac{2\pi}{\underline{b}_0} \sum_{l\neq 0} \left[\int_{\mathbb{R}} d\underline{k}\, |\hat{\mathbf{f}}(\underline{k})|^2 \sum_m |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)| \right. \\ & \left. \times |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} - \frac{2\pi\, l}{\underline{b}_0})\right)|\right]^{\frac{1}{2}} \\ & \times \left[\int_{\mathbb{R}} d\tilde{\underline{k}}\, |\hat{\mathbf{f}}(\tilde{\underline{k}})|^2 \sum_m |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)| \right. \\ & \left. \times |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} + \frac{2\pi\, l}{\underline{b}_0})\right)|\right]^{\frac{1}{2}}. \end{aligned}$$

And finally,

$$|Rest(\mathbf{f})| \leq \frac{2\pi}{\underline{b}_0} \|\mathbf{f}\|^2 \sum_{l\neq 0} \left[\gamma_{\underline{\lambda},\mu}\left(\frac{2\pi}{\underline{b}_0}\underline{k}\right) \gamma_{\underline{\lambda},\mu}\left(-\frac{2\pi}{\underline{b}_0}\underline{k}\right)\right]^{\frac{1}{2}}, \quad (3.46)$$

where,

$$\gamma_{\underline{\lambda},\mu}(\underline{\xi}) = \sup_{\underline{k}\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|\, |\hat{\Psi}\left(a_0^m(\underline{k}+\mu\underline{\lambda}^{-1}) - (\underline{\lambda}^{-1} - \underline{\xi})\right)|.$$

Putting (3.44) and (3.46) together, we get

$$\inf_{\mathbf{f}\in\mathcal{H},\mathbf{f}\neq 0} \|\mathbf{f}\|^{-2} \sum_{m,n} |\langle \Psi_{m,n} \,|\, \mathbf{f} \rangle|^2 \geq \frac{2\pi}{\underline{b}_0} \left\{ \operatorname*{essinf}_{\underline{k}\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|^2 \right.$$

$$\left. - \sum_{l\neq 0} \left[\gamma_{\underline{\lambda},\mu}\left(\frac{2\pi}{\underline{b}_0} l\right) \gamma_{\underline{\lambda},\mu}\left(-\frac{2\pi}{\underline{b}_0} l\right) \right]^{\frac{1}{2}} \right\} \quad 3.47$$

and

$$\sup_{\mathbf{f}\in\mathcal{H},\mathbf{f}\neq 0} \|\mathbf{f}\|^{-2} \sum_{m,n} |\langle \Psi_{m,n} \,|\, \mathbf{f} \rangle|^2 \leq \frac{2\pi}{\underline{b}_0} \left\{ \sup_{\underline{k}\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |\hat{\Psi}\left(a_0^m(\underline{k} + \mu\underline{\lambda}^{-1}) - \underline{\lambda}^{-1}\right)|^2 \right.$$

$$\left. + \sum_{l\neq 0} \left[\gamma_{\underline{\lambda},\mu}\left(\frac{2\pi}{\underline{b}_0} l\right) \gamma_{\underline{\lambda},\mu}\left(-\frac{2\pi}{\underline{b}_0} l\right) \right]^{\frac{1}{2}} \right\}. \quad (3.48)$$

If the right-hand sides of (3.47) and (3.48) are strictly positive and bounded, then the $\Psi_{m,n}$ constitute a frame, and (3.47) gives a lower bound for $A(\underline{\lambda}, \mu)$, (3.48) an upper bound for $B(\underline{\lambda}, \mu)$.

The limiting conditions ($\underline{\lambda} = 0, \mu = 1$) and ($\underline{\lambda} \to 0, \mu = 0$) give the usual frame bounds for the *Gabor* and *wavelet frames* respectively.

Chapter 4

Diffusion Filters and Correspondance with Wavelets

4.1 Introduction

The well known approach of diffusion, starting with Perona-Malik [62] and extended by Weickert and many others, see the references [11, 59, 72, 73], have attracted the attention of Mathematicians and people working in image processing domain. The Mathematicians enjoyed to see the application of PDEs in image processing and for image processing people, it is an another technique for image enhancement / restoration. Since long, wavelet shrinkage was independently used for the same purpose, but recently a beautiful connection has been established between wavelet shrinkage and nonlinear diffusion [72].

In this chapter, we focus mainly on the techniques of diffusion based on partial differential equations (PDE's). We have discussed the detailed account of many different diffusion filter models given especially by Perona-Malik and Weickert. The discretization scheme and the computational results are also given. In the last section, we have given the wavelet shrinkage scheme and the correspondence between diffusion and shrinkage functions. The mathematical justification of this approach, was the observation due to Koenderink [53], where the convolution of an image with the Gaussian kernel is equivalent to the solution of linear diffusion problem with image as an initial condition. The next, important step in the development of this theory was introduced by Perona and Malik [62] in case of nonlinear diffusion model with a more accurate edge detection.

The Perona and Malik model was later axiomatized [3], regularized [11] and modified [4] by Lions and Morel et al. The important contribution to diffusion filter theory is also the work of Weickert [72], where he proposed an extension of the model presented in [11] to the anisotropic case. The main idea of the nonlinear diffusion filtering is based upon adaption of the diffusion process to the image structure. Such an approach allows us to preserve or even enhance edges and at the same time to smooth regions of an image, which do not contain any important information.

The discretization approach used in most of the models mentioned above are based on the finite differences. A detailed finite volume discretization of the afore mentioned models are proposed in Muszkieta [60].

The wavelet shrinkage scheme of Donoho et al. [29] has been very effective in denoising and image enhancement. In a recent paper, Mrázek et al. [59] have successfully established a connection between the shift invariant Haar wavelet shrinkage on onehand and nonlinear diffusion on the other.

4.2 An overview of diffusion process

The diffusion is a physical process that equilibrates concentration differences without creating or destroying mass. This physical observation can easily be described in a mathematical formulation.

The equilibration property is expressed by Fick's law as,

$$j = -D.\nabla u. \tag{4.1}$$

This equation states that a concentration gradient ∇u causes a flux j which aims to compensate for this gradient. The relation between ∇u and j is described by the diffusion tensor D, a positive definite symmetric matrix. The case where j and ∇u are parallel is called isotropic. In this case we may replace the diffusion tensor by a positive scalar valued diffusivity d. In, anisotropic case, j and ∇u are not parallel.

The observation that diffusion does not transport mass without destroying it or creating new mass is expressed by the continuity equation

$$\partial_t u = -\mathrm{div} j \tag{4.2}$$

where t denotes the time .
If we plug Fick's law into the continuity equation we end up with the diffusion equation,

$$\partial_t u = \mathrm{div}(D.\nabla u). \tag{4.3}$$

This equation appears in many physical transport processes. In the context of heat transfer it is called heat equation. In image processing we may identify the concentration with the gray value at a certain location. If the diffusion tensor is constant over the whole image domain, one speaks of homogeneous diffusion and a space dependent filtering is called inhomogeneous. Generally, homogeneous filtering is named isotropic and inhomogeneous blurring is called anisotropic.

4.3 Linear diffusion filter

The simplest and best investigated PDE in image processing is the parabolic linear diffusion equation of the form

$$\frac{\partial u}{\partial t} = \Delta u(x,t), \quad (x,t) \in \mathbb{R}^2 \times (0,\infty) \tag{4.4}$$

with initial condition $u(x,0) = f(x)$ for any $x \in \mathbb{R}^2$. In fact, f is primarily defined only on the domain $\Omega \subset \mathbb{R}^2$, but by symmetry and then periodicity we can extend it to $\mathbb{R}^2$. This method of extension is typical in image processing.

The underlying idea to apply equation (4.4) in image processing comes from the early work of Koenderink [53], who noticed that the convolution of the image f with the Gaussian kernel, defined by

$$G_\sigma(x,y) := \frac{1}{2\pi\sigma^2}\exp\left(-\frac{|x-y|^2}{2\sigma^2}\right) \tag{4.5}$$

is equivalent to the solution u of problem (4.4) for $t = \frac{1}{2}\sigma^2$, that is,

$$u(x,t) = (G_{\sqrt{2t}} * f)(x) = \int_{\mathbb{R}^2} G_{\sqrt{2t}}(x,y)f(y)dy.$$

The above formula gives the correspondence between the time t and the scale parameter σ of the Gaussian kernel G_σ.

The linear diffusion filter has one serious disadvantage. As a matter of fact, it smoothens an image but at the same time blurs important features such as edges, making it difficult to identify on the next stage of image analysis, namely segmentation see the figure 4.2. To overcome this problem one should consider a nonlinear filter, adapted to the local image structure.

4.4 Nonlinear isotropic diffusion filter

4.4.1 The Perona and Malik model

For the first time a nonlinear diffusion filter was introduced by Perona and Malik [62]. They proposed to replace (4.4) by a nonlinear diffusion equation with homogeneous Neumann conditions on the boundary $\partial\Omega$ and solve the following problem in order to obtain the smoothened version of the initial image f

$$\begin{cases} \frac{\partial u}{\partial t} = \nabla.(d(|\nabla u|^2)\nabla u), & (x,t) \in \Omega \times (0,T] \\ u(x,0) = f, & x \in \Omega \\ \frac{\partial u}{\partial n} = 0, & (x,t) \in \Omega \times (0,T]. \end{cases} \tag{4.6}$$

In the first equation, the diffusivity d is positive, monotonically decreasing function, defined in a way, such that the smoothing of image is conditional and depends on its structure. If $|\nabla u|^2$ is large, then the diffusion is low and therefore exact location of edges is kept. If $|\nabla u|^2$ is small, then the diffusion tends to smooth around x. Of course, there exist several possible choices for diffusivity d. As an example, consider the following definition

$$d(s) := \frac{1}{1 + s/\mu} \tag{4.7}$$

where the parameter $\mu > 0$ plays the role of threshold.

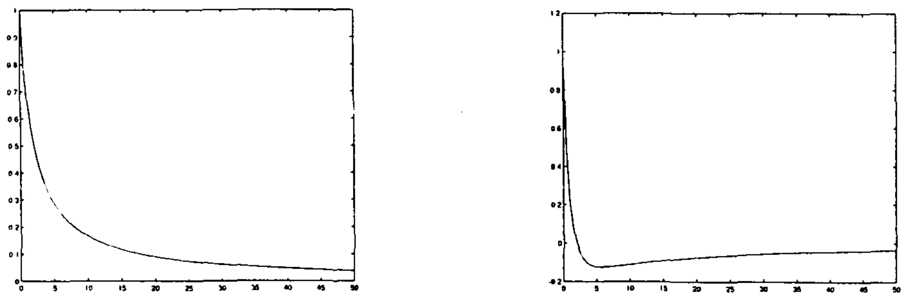

Figure 4.1: (a) Left: Plot of the diffusivity $d(s) = \frac{1}{1+s/\mu}$ with threshold $\mu = 2$ and (b) Right: Plot of the function $b(s) = d(s) + 2sd'(s)$.

However, the Perona and Malik model has several practical and theoretical difficulties. If the image is noisy, then the noise would introduce very large oscillations of the gradient ∇u. Therefore, the adaptive smoothing introduced by the model (4.6) would not give good results, since all these noise edges will be kept. The second difficulty arises from the fact that we obtain a backward diffusion equation for $|\nabla u|^2 > \mu$, which is a classical example of the ill-posed problem. In practice, it implies that very similar images can give divergent solutions and therefore different edges.

Let us explain the second problem in detail. For that, let $\eta = \nabla u/|\nabla u|$ and $\nu = \nabla u^{\perp}/|\nabla u|$ be a vector parallel and perpendicular, respectively, to the gradient ∇u and let us decompose the divergence operator in (4.6) using direction η and ν. We have

$$\nabla.(d(|\nabla u|^2)\nabla u) = d(|\nabla u|^2)\Delta u + 2d'(|\nabla u|^2)\langle \nabla u, \nabla^2 u \nabla u\rangle \tag{4.8}$$

where the expression $\langle \nabla u, \nabla^2 u \nabla u\rangle$ is nothing but the second order derivative of the function u in the gradient ∇u direction.
From the other side, we have

$$\frac{\partial^2 u}{\partial \eta^2} = \langle \eta, \nabla^2 u \eta\rangle = \frac{1}{|\nabla u|^2}\left[\left(\frac{\partial u}{\partial x_1}\right)^2 \frac{\partial^2 u}{\partial x_1^2} + 2\frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_2}\frac{\partial^2 u}{\partial x_1 \partial x_2} + \left(\frac{\partial u}{\partial x_2}\right)^2 \frac{\partial^2 u}{\partial x_2^2}\right]$$

and

$$\frac{\partial^2 u}{\partial \nu^2} = \langle \nu, \nabla^2 u \nu\rangle = \frac{1}{|\nabla u|^2}\left[\left(\frac{\partial u}{\partial x_2}\right)^2 \frac{\partial^2 u}{\partial x_1^2} - 2\frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_2}\frac{\partial^2 u}{\partial x_1 \partial x_2} + \left(\frac{\partial u}{\partial x_1}\right)^2 \frac{\partial^2 u}{\partial x_2^2}\right].$$

Thus, the first equation in (4.6) may be written as follows

$$\frac{\partial u}{\partial t} = d(|\nabla u|^2)\frac{\partial^2 u}{\partial \nu^2} + b(|\nabla u|^2)\frac{\partial^2 u}{\partial \eta^2} \tag{4.9}$$

where $b(s) = d(s) + 2sd'(s)$. Thus, we can interpret equation (4.9) as the sum of a diffusion in the η and ν directions, with function d and b acting as weighting coefficients.

Definition 4.4.1 The partial differential operator $\frac{\partial}{\partial t} + L$, where

$$Lu = -\sum_{i,j=1}^{N} a_{ij}(x,t)\frac{\partial^2 u}{\partial x_i x_j} + \sum_{i=1}^{N} b_i(x,t)\frac{\partial u}{\partial x_i} + c(x,y)u$$

is parabolic for given coefficients $a_{ij}, b_i,\ c\ (i,j = 1, 2, ..., N)$ if and only if there exist constant $C > 0$, such that

$$\sum_{i,j=1}^{N} a_{ij}(x,t)\xi_i\xi_j \geq C|\xi|^2,$$

for all $(x,t) \in \Omega \times (0,T]$ and $\xi \in \mathbb{R}^N$.

We observe that equation (4.9) may be written as follows

$$\frac{\partial u}{\partial t} = a_{11}(|\nabla u|^2)\frac{\partial^2 u}{\partial x_1^2} + 2a_{12}(|\nabla u|^2)\frac{\partial^2 u}{\partial x_1 \partial x_2} + a_{22}(|\nabla u|^2)\frac{\partial^2 u}{\partial x_2^2} \tag{4.10}$$

with

$$a_{11}(|\nabla u|^2) = 2\left(\frac{\partial u}{\partial x_1}\right)^2 d'(|\nabla u|^2) + d(|\nabla u|^2)$$

$$a_{12}(|\nabla u|^2) = 2\frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_2} d'(|\nabla u|^2)$$

$$a_{22}(|\nabla u|^2) = 2\left(\frac{\partial u}{\partial x_2}\right)^2 d'(|\nabla u|^2) + d(|\nabla u|^2).$$

According to Definition 4.4.1, equation (4.10) is parabolic if and only if

$$\sum_{i=1,2} a_{ij}(|\nabla u(x,t)|^2)\xi_i\xi_j > 0,$$

for all $(x, t) \in \Omega \times (0, T]$ and all $\xi \in \mathbb{R}^2$.
An easy calculation shows that this condition reduces to the single inequality

$$b(s) > 0.$$

Let us know examine (4.6) with the diffusivity d defined as in (4.7).
We have

$$d'(s) = \frac{-1}{\mu(1 + s/\mu)^2}$$

and

$$b(s) = d(s) + 2sd'(s) = \frac{\mu(\mu - s)}{(\mu + s)^2}.$$

Therefore, we get $b(|\nabla u|^2) \leq 0$ for $|\nabla u|^2 \geq \mu$. This implies that the model (4.6) with diffusivity d defined in (4.7) fulfill our expections. It is a backward in the direction perpendicular to ∇u, allowing us to sharpen edges.

4.4.2 Regularization of the Perona and Malik model

One way to deal with an ill-posed problem like (4.6) is to introduce regularization which would make the problem well-posed. Then, by reducing the amount of regularization and observing the behavior of the solution of the regularized problem, we can obtain information about the initial one. In the first time, the method to regularize the Perona and Malik problem was proposed by Catté et al. [11]. They suggested to replace the gradient ∇u in the diffusivity $d(|\nabla u|^2)$ by its estimate

$\nabla u_\sigma := \nabla G_\sigma * u$, where G_σ is the Gaussian kernel as defined in (4.5). They have also proven that this slight change is sufficient to ensure existence and uniqueness of the solution to the problem (4.6). This result gives the following theorem

Theorem 4.4.2 Let $d : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}_+$ be smooth, decreasing with $d(0) = 1, \lim_{s\to\infty} d(s) = 0$. If $f \in L^2(\Omega)$, then there exists an unique function $u(x,t) \in C([0,T], L^2(\Omega)) \cap L^2((0,T), H^1(\Omega))$ satisfying in the distributional sense

$$\begin{cases} \frac{\partial u}{\partial t} = \nabla.(d(|\nabla u_\sigma|^2)\nabla u), & (x,t) \in \Omega \times (0,T] \\ u(x,0) = f, & x \in \Omega \\ \frac{\partial u}{\partial n} = 0, & (x,t) \in \partial\Omega \times (0,T]. \end{cases} \tag{4.11}$$

Moreover, $|u|_{L^\infty((0,T),L^2(\Omega))} \leq |f|_{L^2(\Omega)}$ and $u \in C^\infty((0,T) \times \overline{\Omega})$.

4.5 Nonlinear anisotropic diffusion filter

4.5.1 The Weickert model

Despite the advantages of the isotropic diffusion filter, there is still one imperfection: when the diffusion process is stopped near the boundary of an object, it preserves the edges but also leaves a noise at these positions. To avoid this effect, Weickert [72] suggested to modify the diffusion operator so that it diffuses more in direction parallel to edges and less in the perpendicular one. In order to filter an image, he proposed to consider the following problem

$$\begin{cases} \frac{\partial u}{\partial t} = \nabla.(D(S_\rho(\nabla u_\sigma))\nabla u), & (x,t) \in \Omega \times (0,T] \\ u(x,0) = f, & x \in \Omega \\ \langle D(S_\rho(\nabla u_\sigma))\nabla u, n\rangle = 0, & (x,t) \in \partial\Omega \times (0,T]. \end{cases} \tag{4.12}$$

where $D(S_\rho(\nabla u_\sigma))$ is symmetric, positive semidefinite matrix, called the diffusion tensor and it is constructed in the following way.

To avoid false detections of edges due to the presence of noise, we first convolve u with the Gaussian kernel G_σ and calculate the matrix

$$S_0(\nabla u_\sigma) := \nabla u_\sigma^T \nabla u_\sigma. \tag{4.13}$$

This matrix possesses an orthogonal basis composed of eigenvectors v_1, v_2 with $v_1 \parallel \nabla u_\sigma$ and $v_2 \perp \nabla u_\sigma$. The corresponding eigenvalues are equal to $|\nabla u_\sigma|^2$ and

0, respectively, and give contrast in direction of eigenvectors.

In next step, the local information is averaged by convolving S_0 component wise with the Gaussian kernel G_ρ. As a result we obtain symmetric, positive semidefinite matrix

$$S_\rho(\nabla u_\sigma)) := G_\rho * S_0(\nabla u_\sigma) := \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \tag{4.14}$$

The matrix S_ρ is called the structure tensor and possesses orthonormal eigenvectors v_1, v_2 with v_1 parallel to

$$\begin{bmatrix} 2s_{12} \\ \\ s_{22} - s_{11} + \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \end{bmatrix}.$$

The corresponding eigenvalues are given by

$$\mu_1 = \frac{1}{2}\left[s_{11} + s_{22} + \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \right]$$

and

$$\mu_2 = \frac{1}{2}\left[s_{11} + s_{22} - \sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2} \right].$$

This diffusion tensor $D(S_\rho(\nabla u_\sigma))$ should have the same set of eigenvectors as the structure tensor S_ρ, and the choice of corresponding eigenvalues should depend on the desired goal of the filter.

Diffusion tensor. Let us now describe how to define the diffusion tensor $D(S_\rho(\nabla u_\sigma))$ such that the anisotropic filter (4.12) could be applied to particular problems.

Since the eigenvectors of the diffusion tensor should reflect the local image structure, one should choose the same orthonormal basis of eigenvectors that one gets from the structure tensor S_ρ. Following Weickert [72], we introduce here two possible choices of eigenvalues λ_1 and λ_2 of $D(S_\rho(\nabla u_\sigma))$.

Edge-enhancing anisotropic diffusion. Suppose one wants to smooth an image within some region and at the same time preserve edges, then one should reduce the diffusivity λ_1 perpendicular to edges more if the contrast given by the greatest eigenvalue μ_1 of the structure tensor S_ρ is large. This behavior may be accomplished by the following choice of eigenvalues,

$$\lambda_1 := \begin{cases} 1, & \text{if } \mu_1 \leq 0 \\ \\ 1 - \exp\left(\frac{-3.315}{\mu_1^4}\right), & \text{if } \mu_1 > 0 \end{cases}$$

$$\lambda_2 := 1.$$

Coherence-enhancing anisotropic diffusion. If one wants to enhance coherent structures, then one should perform smoothening, preferably along the coherence direction v_2 with diffusivity λ_2, which is increasing with respect to the coherence $(\mu_1 - \mu_2)^2$. This may be achieved by the following choice of eigenvalues of the diffusion tensor,

$$\lambda_1 := \alpha$$

$$\lambda_2 := \begin{cases} \alpha, & \text{if } \mu_1 = \mu_2 \\ \alpha + (1-\alpha)\exp\left(\frac{-1}{(\mu_1-\mu_2)^2}\right), & \text{otherwise} \end{cases}$$

where $\alpha \in (0,1)$ is a small positive parameter which keeps the diffusion tensor $D(S_\rho(\nabla u_\sigma))$ uniformly positive definite.

4.6 Explicit discretization scheme

When applied to discrete signals, the partial differential equation (4.6) has to be discretized. In this section, we focus on explicit finite difference schemes. Substituting the spatial partial derivatives in (4.6) by finite differences (with the assumption of unit distance between neighboring pixels), and employing explicit discretization in time, an explicit 1-D scheme for nonlinear diffusion can be written in the form

$$\frac{u_i^{k+1} - u_i^k}{\tau} = d(|u_{i+1}^k - u_i^k|)(u_{i+1}^k - u_i^k) - d(|u_i^k - u_{i-1}^k|)(u_i^k - u_{i-1}^k),$$

where τ is the time step size and the upper index k denotes the approximate solution at time $k\tau$. Separating the unknown u_i^{k+1} on one side, we obtain

$$u_i^{k+1} = u_i^k - \tau d(|u_i^k - u_{i+1}^k|)(u_i^k - u_{i+1}^k) + \tau d(|u_{i-1}^k - u_i^k|)(u_{i-1}^k - u_i^k). \tag{4.15}$$

The initial condition reads $u_i^0 = f_i$ for all i.

4.7 Computational result

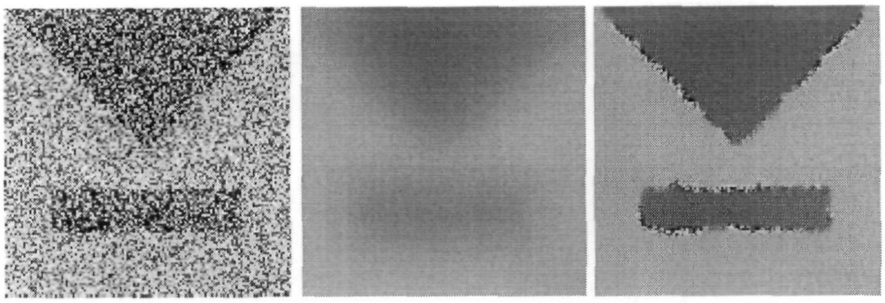

Figure 4.2: Restoration properties of diffusion filters. (a) left: Noisy image, (b) middle: Image is processed for linear diffusion (c) right : Nonlinear isotropic diffusion

4.8 Wavelet shrinkage

4.8.1 Basic concept

The discrete wavelet transform represents a one dimensional signal in terms of shifted versions of a dilated lowpass scaling function φ, and shifted and dilated versions of a bandpass wavelet function ψ. In case of orthonormal wavelets, this gives

$$f = \sum_{i \in \mathbb{Z}} \langle f, \varphi_i^n \rangle \varphi_i^n + \sum_{j=-\infty}^{n} \sum_{i \in \mathbb{Z}} \langle f, \psi_i^j \rangle \psi_i^j. \tag{4.16}$$

Here we have taken the convention, $\psi_i^j(s) := 2^{-j/2}\psi(2^{-j}s - i)$ and $\langle .,. \rangle$ denotes the inner product in $L^2(\mathbb{R})$. If the measurement f is corrupted by moderate white Gaussian noise, then this noise is contained to a small amount in all wavelet coefficients $\langle f, \psi_i^j \rangle$, while the original signal is in general determined by a few significant wavelet coefficients [55]. Therefore, wavelet shrinkage attempts to eliminate noise from the wavelet coefficients by the following three-step procedure:

1. Analysis: transform the noisy data f to the wavelet coefficients $d_i^j = \langle f, \psi_i^j \rangle$ and scaling function coefficients $c_i^n = \langle f, \varphi_i^n \rangle$.

2. Shrinkage: apply a shrinkage function S_θ with a threshold parameter θ to the wavelet coefficients, i.e., $S_\theta(d_i^j) = S_\theta(\langle f, \psi_i^j \rangle)$.

3. Synthesis: reconstruct the denoised version u of f from the shrunken wavelet coefficients.

$$u := \sum_{i\in\mathbb{Z}} \langle f, \varphi_i^n\rangle \varphi_i^n + \sum_{j=-\infty}^{n} \sum_{i\in\mathbb{Z}} S_\theta(\langle f, \psi_i^j\rangle)\psi_i^j.$$

In this chapter, we restrict our attention to Haar wavelets, well suited for piecewise constant signals with discontinuities. The Haar wavelet and scaling functions are given respectively by

$$\psi(x) = 1_{[0,\frac{1}{2})} - 1_{[\frac{1}{2},1)}, \tag{4.17}$$

$$\phi(x) = 1_{[0,1)}, \tag{4.18}$$

where $1_{[a,b)}$ denotes the characteristic function, equal to 1 on $[a, b)$ and zero everywhere else. Using the so-called "two-scale relation" of the wavelet and its scaling function, the coefficients c_i^j and d_i^j at higher level j can be computed from the coefficients c_i^{j-1} at lower level $j - 1$ and conversely, i.e.,

$$c_i^j = \frac{c_{2i}^{j-1} + c_{2i+1}^{j-1}}{\sqrt{2}}, \quad d_i^j = \frac{c_{2i}^{j-1} - c_{2i+1}^{j-1}}{\sqrt{2}}, \tag{4.19}$$

and

$$c_{2i}^{j-1} = \frac{c_i^j + d_i^j}{\sqrt{2}}, \quad c_{2i+1}^{j-1} = \frac{c_i^j - d_i^j}{\sqrt{2}}. \tag{4.20}$$

This results in a fast algorithm for the analysis and synthesis steps. Various shrinkage functions leading to qualitatively different denoised functions u were considered in literature, e.g.,

A. Linear shrinkage: $S(x) = \lambda x \quad (\lambda \in [0, 1]),$

B. Soft shrinkage [28]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x - \theta \operatorname{sgn}(x) & |x| > \theta. \end{cases}$

C. Garrote shrinkage [37]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x - \frac{\theta^2}{x} & |x| > \theta. \end{cases}$

D. Firm shrinkage [38]: $S_{\theta_1,\theta_2}(x) = \begin{cases} 0 & |x| \le \theta_1, \\ \operatorname{sgn}(x)\frac{\theta_2(|x|-\theta_1)}{\theta_1-\theta_2} & \theta_1 < |x| \le \theta_2, \\ x & \theta_2 < |x|. \end{cases}$

E. Hard shrinkage [55]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x & |x| > \theta. \end{cases}$

4.8.2 Discrete translation invariant scheme

In practice one deals with discrete signals $\mathbf{f} = (f_i)_{i=0}^{N-1}$, where for simplicity N is a power of 2. The Haar wavelet shrinkage starts by setting $c_i^0 = f_i$ and proceeds by analysis, shrinkage and synthesis. Let us just consider a single wavelet decomposition level, i.e., we set $n = 1$. Then, using the convention that $c_i = c_i^1$ and $d_i = d_i^1$, we can drop the superscripts $j = 0$ and $j = 1$. By (4.19) and (4.20), Haar wavelet shrinkage on one level produces the signal $\mathbf{u}^+ = (u_i^+)_{i=0}^{N-1}$ with coefficients

$$u_{2i}^+ = \frac{c_i + S_\theta(d_i)}{\sqrt{2}} = \frac{f_{2i} + f_{2i+1}}{2} + \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i} - f_{2i+1}}{\sqrt{2}}\right), \tag{4.21}$$

$$u_{2i+1}^+ = \frac{c_i - S_\theta(d_i)}{\sqrt{2}} = \frac{f_{2i} + f_{2i+1}}{2} - \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i} - f_{2i+1}}{\sqrt{2}}\right). \tag{4.22}$$

Note that the single Haar wavelet shrinkage step (4.21)-(4.22) decouples the input signal into successive pixel pairs: the pixel at position $2i - 1$ has no direct connection to its neighbour at position $2i$, and the procedure is not invariant to translation of the input signal. To overcome this problem, Coifman and Donoho [20] introduced the so called cycle spinning: the input signal is shifted, denoised using wavelet shrinkage, shifted back, and the results of all such shifts are averaged. For our single decomposition level, we need only one additional shift to acquire translation invariance. The shifted Haar wavelet shrinkage yields the signal $\mathbf{u}^- = (u_i^-)_{i=0}^{N-1}$ with coefficients,

$$u_{2i-1}^- = \frac{f_{2i-1} + f_{2i}}{2} + \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i-1} - f_{2i}}{\sqrt{2}}\right),$$

$$u_{2i}^- = \frac{f_{2i-1} + f_{2i}}{2} - \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i-1} - f_{2i}}{\sqrt{2}}\right).$$

Averaging the shifted results, one cycle of shift invariant Haar wavelet shrinkage can be summarized into

$$\begin{aligned} u_i &= \frac{u_i^- + u_i^+}{2} \\ &= \frac{f_{i-1} + 2f_i + f_{i+1}}{4} + \frac{1}{2\sqrt{2}} S_\theta\left(\frac{f_i - f_{i+1}}{\sqrt{2}}\right) - \frac{1}{2\sqrt{2}} S_\theta\left(\frac{f_{i-1} - f_i}{\sqrt{2}}\right) \end{aligned} \tag{4.23}$$

4.9 Correspondence between diffusion and shrinkage functions

4.9.1 Basic considerations

In order to derive the relation between the explicit diffusion scheme and translation invariant Haar wavelet shrinkage, we rewrite the first iteration step in (4.15) using the initial condition $u_i^0 = f_i$ and the simplified $u_i^1 = u_i$ as

$$u_i = \frac{f_{i-1} + 2f_i + f_{i+1}}{4} + \frac{f_i - f_{i+1}}{4} - \frac{f_{i-1} - f_i}{4}$$

$$
\begin{aligned}
& - \tau d(|f_i - f_{i+1}|)(f_i - f_{i+1}) + \tau d(|f_{i-1} - f_i|)(f_{i-1} - f_i) \\
= & \frac{f_{i-1} + 2f_i + f_{i+1}}{4} \\
& +(f_i - f_{i+1}) \left(\frac{1}{4} - \tau d(|f_i - f_{i+1}|) \right) \\
& -(f_{i-1} - f_i) \left(\frac{1}{4} - \tau d(|f_{i-1} - f_i|) \right) . \qquad (4.24)
\end{aligned}
$$

This coincides with (4.23) if and only if

$$\frac{1}{2\sqrt{2}} S_\theta \left(\frac{x}{\sqrt{2}} \right) = x \left(\frac{1}{4} - \tau d(|x|) \right) . \qquad (4.25)$$

Equation (4.25) relates the shrinkage function S_θ of wavelet denoising to the diffusivity d of nonlinear diffusion. Provided that relation (4.25) holds true, a single step of wavelet shrinkage is equivalent to a single step of explicitly discretized nonlinear diffusion. The following two formulae are derived from (4.25) and can be used to obtain a shrinkage function S_θ from a diffusivity d, or vice versa.

$$S_\theta(x) = x(1 - 4\tau d(|\sqrt{2}x|)), \qquad (4.26)$$

$$d(|x|) = \frac{1}{4\tau} - \frac{\sqrt{2}}{4\tau x} S_\theta \left(\frac{x}{\sqrt{2}} \right) . \qquad (4.27)$$

4.9.2 From diffusivities to shrinkage functions

Let us now investigate equation (4.26) in detail. The examples from Section 4.8.1 show that typical shrinkage from the literature satisfy

$$S(x) \geq 0 \quad \text{for} \quad x > 0, \qquad (4.28)$$

$$S(x) \leq 0 \quad \text{for} \quad x < 0. \qquad (4.29)$$

One can show that these conditions are responsible for ensuring certain stability properties (so called sign stability) of the shrinkage process. We can now specify the time step size τ in (4.26) such that these two conditions are always satisfied for bounded diffusivities. The diffusivities 1-4 below are bounded by 1.

1. Linear diffusivity [48]: $d(|x|) = 1$,
2. Charbonnier diffusivity [12]: $d(|x|) = \frac{1}{\sqrt{1+\frac{|x|^2}{\lambda^2}}}$,
3. Perona-Malik diffusivity [62]: $d(|x|) = \frac{1}{1+\frac{|x|^2}{\lambda^2}}$,

4. Weickert diffusivity [72]: $$d(|x|) = \begin{cases} 1 & |x| = 0 \\ 1 - \exp\left(\frac{-3.31488}{(|x|/\lambda)^8}\right) & |x| > 0, \end{cases}$$

5. TV diffusivity [5]: $d(|x|) = \frac{1}{|x|}$,

6. BFB diffusivity [52]: $d(|x|) = \frac{1}{|x|^2}$.

In order to ensure that the corresponding shrinkage functions satisfy (4.28)-(4.29), the time step size has to fulfill $\tau \leq 0.25$.

We observe that the linear diffusivity corresponds to the linear shrinkage function

$$S(x) = (1 - 4\tau)x.$$

Nonlinear shrinkage functions such as soft, garrote, firm and hard shrinkage satisfy $S'(0) = 0$, since the goal was to set small wavelet coefficients to zero. In order to derive shrinkage functions that correspond to the bounded nonlinear diffusivities 2-4 and satisfy $S'(0) = 0$ as well, we fix $\tau := 0.25$. Then we obtain the following novel shrinkage functions.

(i) The Charbonnier diffusivity corresponds to the shrinkage function

$$S_\lambda(x) = x\left(1 - \sqrt{\frac{\lambda^2}{\lambda^2 + 2x^2}}\right).$$

(ii) The Perona-Malik diffusivity leads to

$$S_\lambda(x) = \frac{2x^3}{2x^2 + \lambda^2}.$$

(iii) The Weickert diffusivity gives

$$S_\lambda(x) = \begin{cases} 0 & x = 0 \\ x\exp\left(\frac{-0.20718\lambda^8}{x^8}\right) & x \neq 0. \end{cases}$$

4.9.3 From shrinkage functions to diffusivities

Having derived shrinkage functions from nonlinear diffusivities, let us now derive diffusivities from frequently used shrinkage functions. To this end, all we have to do is to plug in the specific shrinkage function into (4.27).
In the case of soft shrinkage, this gives the diffusivity:

$$d(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \theta\sqrt{2}, \\ \frac{\sqrt{2}\theta}{4\tau|x|} & |x| > \theta\sqrt{2}. \end{cases}$$

If we select the time step size τ such that $\theta = 2\sqrt{2}\tau$, we obtain a stabilised TV diffusivity:

$$d(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq 4\tau, \\ \frac{1}{|x|} & |x| > 4\tau. \end{cases}$$

In the same way one can show that garrote shrinkage leads to a stabilised BFB diffusivity for $\theta = \sqrt{2\tau}$:

$$d(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq 2\sqrt{\tau}, \\ \frac{1}{|x|^2} & |x| > 2\sqrt{\tau}. \end{cases}$$

Firm shrinkage yields a diffusivity that degenerates to 0 for sufficiently large gradients:

$$d(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \sqrt{2}\theta_1, \\ \frac{\theta_1}{4\tau(\theta_2-\theta_1)}\left(\frac{\sqrt{2}\theta_2}{|x|} - 1\right) & \sqrt{2}\theta_1 < |x| \leq \sqrt{2}\theta_2, \\ 0 & |x| > \sqrt{2}\theta_2. \end{cases}$$

Another diffusivity that degenerates to 0 can be derived from hard shrinkage:

$$d(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \sqrt{2}\theta, \\ 0 & |x| > \sqrt{2}\theta. \end{cases}$$

Chapter 5

Sharp Operator based Diffusion Filter

5.1 Introduction

In this Chapter we have introduced a new difusion filter based on Sharp Operator. We have studied the Hardy-Littlewood maximal function and the sharp operator to measure the oscillatory behaviour of images. With the sharp operator we measure the oscillation in the neighbourhood of pixel of an images.

The maximal function was introduced by Hardy and Littlewood [44] to solve a problem in the theory of functions of complex variable. Based on this idea John and Nirenberg [49] introduced the concept of Bounded Mean Oscillation (BMO) functions. Fefferman and Stein [35] introduced the sharp function (denoted by $f^{\#}$) and found that a function $f \in BMO$ is equivalent with $f^{\#} \in L_{\infty}$. The theory of H^p spaces (Hardy Spaces) received impetus from the work of Fefferman and Stein. Their work resulted in the identification of the dual of H^1 with BMO. The idea of applying the sharp operator to measure the oscillation and classification of images was proposed by Ahmad and Siddiqi [1] where it is used in choosing a proper compression technique.

In Section 5.2, we have described the theoretical results of the maximal function, BMO and the sharp function. In Section 5.3, we have defined the diffusion filters. In Section 5.5, we have presented our computational results.

5.2 Theoretical results

The Hardy-Littlewood maximal function was developed to solve a problem in the theory of functions of complex variable. The analogue for integrals, which is required for the function theoretic applications, is also derived by Hardy and Littlewood [44].

Definition 5.2.1 Let $\mathbb{R}^n$ be the n-dimensional Euclidean space and $f(x)$ be a real valued measurable function on $\mathbb{R}^n$. For such a function f on $\mathbb{R}^n$ its Hardy-Littlewood maximal function is defined by the formula

$$Mf(x) = \sup_Q \left\{ \frac{1}{\lambda(Q)} \int_Q |f(y)| dy \ : \ Q \subset \mathbb{R}^n, x \in Q \right\}, \tag{5.1}$$

where the supremum ranges over all finite cubes Q in $\mathbb{R}^n$ and $\lambda(Q)$ is the Lebesgue measure of Q.

The function $Mf(x)$ has the following properties:

(i) $0 \leq Mf(x) \leq \infty$

(ii) $M(f+g)(x) \leq Mf(x) + Mg(x)$

(iii) $M(\alpha f)(x) = |\alpha| Mf(x)$

where f, g are measurable functions on $\mathbb{R}^n$ and α is some scalar quantity.

It is easy to find a function whose maximal function is un-bounded.

Example 5.2.2 For $f(x) = |x|^t$ with $t > 0$, we get $Mf(x) = \infty$ for each $x \in \mathbb{R}$.

Now we state a Hardy-Littlewood maximal theorem.

Theorem 5.2.3 For each function $f \in L_1(\mathbb{R}^n)$ we have

$$\lambda(\{x : Mf(x) > t\}) \leq 6^n t^{-1} \|f\|_1, \ \ t > 0 \, .$$

Proof. See [77], page 142.

An interesting application of the maximal theorem is a version of the Lebesgue differentiation theorem.

Theorem 5.2.4 (Lebesgue Differentiation Theorem). Let $f \in L_1(\mathbb{R}^n)$. For almost all $x \in \mathbb{R}^n$ and for every decreasing sequence of cubes $(Q_j)_{j=1}^{\infty}$, such that $\cap_{j=1}^{\infty} Q_j = \{x\}$, we have

$$\lim_{j \to \infty} \frac{1}{\lambda(Q_j)} \int_{Q_j} f(y) dy = f(x). \tag{5.2}$$

Proof. See [68], page 81.

The space BMO, i.e. bounded mean oscillation of functions is introduced by John and Nirenberg [49].

Definition 5.2.5 A measurable function f on $\mathbb{R}^n$ has bounded p-mean oscillation, $1 \leq p < \infty$, if

$$\|f\|_{BMO_p} = \sup_Q \left(\frac{1}{\lambda(Q)} \int_Q |f(x) - f_Q|^p dx \right)^{\frac{1}{p}} < \infty, \tag{5.3}$$

where the supremum ranges over all finite cubes Q in $\mathbb{R}^n$ and $f_Q = \frac{1}{\lambda(Q)} \int_Q f(x)dx$ is the mean value of the function f on the cube Q.

The set of all functions of bounded p-mean oscillation is denoted by $BMO_p(\mathbb{R}^n)$. $\|f\|_{BMO_p}$ is "almost" a norm since it has the following properties.

(i) $\|f + g\|_{BMO_p} \leq \|f\|_{BMO_p} + \|g\|_{BMO_p}$

(ii) $\|\alpha f\|_{BMO_p} = |\alpha| \cdot \|f\|_{BMO_p}$

(iii) $\|f\|_{BMO_p} = 0$ if and only if $f=$ constant almost everywhere,

where f, g are measurable functions on $\mathbb{R}^n$ and α is some scalar quantity.

If we define

$$\|f\|_{BMO'_p} = \left| \int_{\mathbb{R}^n} f(x)dx \right| + \sup_Q \left(\frac{1}{\lambda(Q)} \int_Q |f(x) - f_Q|^p dx \right)^{\frac{1}{p}},$$

we get a norm of f and BMO'_p becomes a Banach Space. On the other way we can say, $\|f\|_{BMO_p}$ becomes a norm if we identify functions which differ by a constant. With this identification $BMO_p(\mathbb{R}^n)$ becomes a normed space, and ultimately a Banach space.

Fefferman and Stein [35] introduced "sharp function" $f^\#$ that mediates between BMO_p and L_p spaces. It is defined as follows.

Definition 5.2.6 Let f is locally integrable function on $\mathbb{R}^n$. The sharp function $f^\#(x)$ is represented by the formula,

$$f^\#(x) = \sup_{Q:x\in Q} \left(\frac{1}{\lambda(Q)} \int_Q |f(y) - f_Q|^p dy \right)^{\frac{1}{p}}. \tag{5.4}$$

Of course, $f \in BMO_p$ is identical with $f^\# \in L_\infty$. It is also observed that there are unbounded functions in $BMO_p(\mathbb{R})$.

Example 5.2.7 The function $f(x) = \ln|x|$ on $\mathbb{R}$ is in $BMO_1(\mathbb{R})$.

After calculation it comes out to be $\|\ln|x|\|_{BMO_1} \leq 2$. So, the un-bounded function $\ln|x|$ is in $BMO_1(\mathbb{R})$.

It is important to note that it does not matter in which L_p norm we measure the oscillation. This is clear from the following corollary.

Corollary 5.2.8 For each p, $1 \leq p < \infty$, there exists a constant C_p such that for each $f \in BMO_p(\mathbb{R}^n)$ we have

$$\|f\|_{BMO_1} \leq \|f\|_{BMO_p} \leq C_p \|f\|_{BMO_1}$$

Proof. See [77], page 156.

In view of the above corollary the spaces $BMO_p(\mathbb{R}^n)$ are equivalent for all p, $1 \leq p < \infty$.

5.3 Diffusion filters

It is clear from the definition of the sharp function that for a pixel z in an almost uniform grey level region in an image, $f^{\#}(z)$ will be very small. However, for the contrast region we get large $f^{\#}(z)$ values.

The idea is to accrue more diffusion in the regions of lower oscillation whereas to preserve the regions of higher oscillation. One of the serious problems of the diffusion model is that it is very sensitive to noise. The noise often introduces very large oscillation in the gradient ∇u, therefore the gradient bases diffusivity, like $d(|\nabla u|)$, possibly miscontrues the true edges and that leads to undesirable diffusion in important regions. Thus a remedy to the deficiency is suggested by introducing a regularization of the model by Gaussian [62].

We have tried to evaluate the performance of sharp operator as a new diffusion entity, defined as

$$d(f^{\#}(x)) = \frac{1}{1 + f^{\#n}(x)/\lambda}, \tag{5.5}$$

where λ is a contrast parameter and can be adjusted.

After regularization of model with the Gaussian we have used sharp function to calculate the diffusivity. Our model can be written as:

$$\frac{\partial u}{\partial t} = \nabla . \left(d \left(f^{\#} u_{\sigma} \right) \nabla u \right). \tag{5.6}$$

Let Ω denotes the open set $(0,1) \times (0,1)$ of $\mathbb{R}^2$, with boundary Γ. We denote $H^k(\Omega), k$ a positive integer, the set of all function $u(x)$ defined in Ω such that u and its distributional derivatives $D^s(u)$ of order $|s| = \sum_{j=1}^n s_j \leq k$ all belong to $L^2(\Omega)$. $H^k(\Omega)$ is a Hilbert space for the norm

$$\|u\|_{H^k(\Omega)} = \left(\sum_{|s| \leq k} \int_\Omega |D^s u(x)|^2 dx \right)^{1/2}.$$

We denote $L^p(0, T, H^k(\Omega))$, the set of all function u, such that for almost every t in $(0, T)$, $u(t)$ belong to $H^k(\Omega)$. $L^p(0, T, H^k(\Omega))$ is a normed space for the norm

$$\|u\|_{L^p(0,T,H^k(\Omega))} = \left(\int_0^T \|u(t)\|^p_{H^k(\Omega)} \right)^{1/p},$$

$p > 1$ and k a positive integer.

We denote $L^\infty(0, T, C^\infty(\Omega))$, the set of all function such that, for almost every t in $(0, T)$, $u(t)$ belong to $C^\infty(\Omega)$. $L^\infty(0, T, C^\infty(\Omega))$ is a normed space for the norm

$$\|u\|_{L^\infty(0,T,C^\infty(\Omega))} = \inf\{C; \|u(t)\|_{C^\infty(\Omega)} \leq C, \text{a.e. on } (0, T)\}.$$

Let $d : R^+ \to R^+$ be a decreasing function with $d(0) = 1$ and $\lim_{t\to\infty} d(t) = 0$.

Theorem 5.3.1 Let $u_0 \in L^2(\Omega)$, then we have a unique function $u(x, t)$ such that $u \in C([0, T]; L^2(\Omega)) \cap L^2(0, T, H^1(\Omega))$, and verifying

$$\begin{cases} \frac{\partial u}{\partial t} - \nabla.(d(f^\# u_\sigma)\nabla u) = 0 \text{ on } (0, T) \times \Omega, \\ \frac{\partial u}{\partial n} = 0 \text{ on } \Gamma \times (0, T], \\ u(0) = u_0, \end{cases} \tag{5.7}$$

where $u_\sigma = G_\sigma * u$ and system is verified in the distributional sense. Moreover, this unique solution is in $C^\infty((0, T) \times \overline{\Omega})$.

Proof. Existence of solution. Here, we show the existence of a weak solution of (5.7) by a classical fixed point theorem of Schauder [36]. We take the space

$$S(0, T) = \left\{ s \in L^2(0, T, H^1(\Omega)),\ \frac{ds}{dt} \in L^2(0, T, (H^1(\Omega))') \right\}.$$

This space is a Hilbert space for the graph norm [54]. Let $s \in S(0, T) \cap L^\infty(0, T, L^2(\Omega))$ such that

$$\|s\|_{L^\infty(0,T,L^2(\Omega))} \leq \|u_0\|_{L^2(\Omega)} \tag{$*$}$$

and consider the problem (X_s):

$$\left\langle \frac{\partial u}{\partial t}(t), v \right\rangle + \int_\Omega d(f^\#(G_\sigma * s)(t))\nabla u(t)\nabla v = 0, \quad \forall\, v \in H^1(\Omega) \text{ a.e. in } [0,T],$$

$$u(0) = u_0.$$

Since $s \in L^\infty(0,T,L^2(\Omega))$ and d, G are C^∞, one can easily deduce that $d(f^\# u_\sigma) \in L^\infty(0,T,C^\infty(\Omega))$. Thus , since d is a decreasing , there exists a constant $\nu > 0$ such that

$$d(f^\#(G_\sigma * s)) \geq \nu \quad \text{a.e. in } (0,T)\times\Omega,$$

where ν depends only on d, G and $\|u_0\|_{L^2(\Omega)}$.

By classical results on parabolic equations [7, 9], we prove that the problem (X_s) as a unique solution $P(s)$ in $S(0,T)$ [2, 9]. Then we can deduce

$$\|P(s)\|_{L^2(0,T,H^1(\Omega))} \leq c_1, \tag{5.8}$$

$$\|P(s)\|_{L^\infty(0,T,L^2(\Omega))} \leq \|u_0\|_{L^2(\Omega)}, \tag{5.9}$$

$$\|P(s)\|_{L^2(0,T,(H^1(\Omega))')} \leq c_2, \tag{5.10}$$

where c_1 and c_2 are constants. These estimates lead us to introduce the subset S_0 of $S(0,T)$ defined by

$$S_0 = \left\{ \begin{array}{r} s \in S(0,T);\, \|s\|_{L^2(0,T,H^1(\Omega))} \leq c_1,\, \|s\|_{L^\infty(0,T,L^2(\Omega))} \leq \|u_0\|_{L^2(\Omega)} \\ \|\frac{ds}{dt}\|_{L^2(0,T,(H^1(\Omega))')} \leq c_2,\, s(0) = u_0 \end{array} \right\}.$$

By (5.8)-(5.10), P is a mapping from S_0 into S_0. Moreover, S_0 is a nonempty, convex and weakly compact subset of $S(0,T)$.

In order to use the Schauder theorem, we need to prove that the mapping u is weakly continuous from S_0 into S_0.

Since $S(0,T)$ is contained in $L^2(0,T,L^2(\Omega))$ with compact inclusion, this will provide u in S_0 such that $u = P(u)$.

Let (s_j) be a sequence in S_0 which converges weakly to some s in S_0 and $u_j = P(s_j)$.

By using (5.8)-(5.10) and classical theorems of compact inclusion (the theorem of Rellich and Kondrachov [9]), the sequence (s_j) of S_0 contains a subsequence (s_j)

such that

$u_j \to u$ weakly in $L^2(0,T,H^1(\Omega))$,

$\frac{du_j}{dt} \to \frac{du}{dt}$ weakly in $L^2(0,T,(H^1(\Omega))')$,

$u_j \to u$ in $L^2(0,T,L^2(\Omega))$ and a.e. in $\Omega \times (0,T)$,

$\frac{\partial u_j}{\partial x_i} \to \frac{\partial u}{\partial x_i}$ weakly in $L^2(0,T,L^2(\Omega))$, $i = 1,2$

$s_j \to s$ in $L^2(0,T,L^2(\Omega))$,

Thus $d(f^{\#}(G * s_j)) \to d(f^{\#}(G * s))$ in $L^2(0,T,L^2(\Omega))$,

$u_j(0) \to u$ in $(H^1(\Omega))'$.

Then we can pass to the limit in the relation (Xs_j), which yield $u = P(s)$. Moreover, by the uniqueness of the solution of (X_s), the whole sequence $u_j = P(s_j)$ converges weakly in $S(0,T)$ to $u = P(s)$.

Uniqueness of the solution. Let $\overline{u}$ and $\hat{u}$ be two solution of (5.7). We have, for almost every t in $[0,T]$,

$$\frac{d\overline{u}}{dt}(t) - \operatorname{div}(\overline{\alpha}(t)\nabla\ \overline{u}(t)) = 0,\ \frac{\partial\ \overline{u}}{\partial n}(t) = 0,\ \overline{u}(0) = u_0 \tag{5.11}$$

$$\frac{d\hat{u}}{dt}(t) - \operatorname{div}(\hat{\alpha}(t)\nabla\ \hat{u}(t)) = 0,\ \frac{\partial\ \hat{u}}{\partial n}(t) = 0,\ \hat{u}(0) = u_0. \tag{5.12}$$

Where $\overline{\alpha}(t) = d(f^{\#}(G_\sigma * \overline{u})(t))$ and $\hat{\alpha}(t) = d(f^{\#}(G_\sigma * \hat{u})(t))$.

By using (5.11) and (5.12), we obtain

$$\begin{aligned}\frac{d}{dt}(\overline{u}(t) - \hat{u}(t)) - \operatorname{div}[\ \overline{\alpha}(t)(\nabla\ \overline{u}(t) - \nabla\ \hat{u}(t)]&\\ = \operatorname{div}[\ (\overline{\alpha}(t) - \hat{\alpha}(t))\nabla\ \hat{u}t].&\end{aligned} \tag{5.13}$$

Now , multiplying (5.13) by $\overline{u}(t) - \hat{u}(t)$ and using (*)

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}(\|\overline{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}) + \nu\|\nabla\ \overline{u}(t) - \nabla\hat{u}(t)\|^2_{L^2(\Omega)}\\ &\leq \|\overline{\alpha}(t) - \hat{\alpha}(t)\|^2_{L^\infty(\Omega)}\|\nabla\hat{u}(t)\|_{L^2(\Omega)}\|\nabla\ \overline{u}(t) - \nabla\hat{u}(t)\|_{L^2(\Omega)}.\end{aligned} \tag{5.14}$$

Since d, G are C^∞, we have

$$\|\overline{\alpha}(t) - \hat{\alpha}(t)\|_{L^\infty(\Omega)} \leq c\|\overline{u}(t) - \hat{u}(t)\|_{L^2(\Omega)}, \tag{5.15}$$

where c is constant.
By using (5.14) and (5.15), we have

$$\frac{1}{2}\frac{d}{dt}(\|\overline{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}) + \nu\|\nabla\ \overline{u}(t) - \nabla\hat{u}(t)\|^2_{L^2(\Omega)}$$

$$\leq \frac{2}{\nu}c\|\overline{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}\|\nabla\hat{u}(t)\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|\nabla\ \overline{u}(t) - \nabla\hat{u}(t)\|^2_{L^2(\Omega)},$$

and so

$$\frac{d}{dt}(\|\overline{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}) \leq \frac{4c}{\nu}\|\overline{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}\|\nabla\hat{u}(t)\|^2_{L^2(\Omega)}. \tag{5.16}$$

Since $\overline{u}(0) = \hat{u}(0) = u_0$, by using (5.16) and Gronwall's Lemma [10] we obtain the desired result.

5.3.1 Discretization step

The finite difference scheme for the model

$$\frac{\partial u}{\partial t} = \nabla.(d(f^{\#}(u_\sigma))\nabla u)$$

is given by,

$$\frac{\partial u}{\partial t} = \partial x_1(d(f^{\#}(u_\sigma))\partial x_1 u) + \partial x_2(d(f^{\#}(u_\sigma))\partial x_2 u).$$

More explicitly, the scheme can be written in the following form:

$$\begin{aligned}\frac{u^{k+1}_{i,j} - u^k_{i,j}}{\tau} &= d(f^{\#k}_{i+1,j})(u^k_{i+1,j} - u^k_{i,j}) - d(f^{\#k}_{i,j})(u^k_{i,j} - u^k_{i-1,j}) \\ &+ d(f^{\#k}_{i,j+1})(u^k_{i,j+1} - u^k_{i,j}) - d(f^{\#k}_{i,j})(u^k_{i,j} - u^k_{i,j-1}),\end{aligned}$$

where $u^k_{i,j}$ denotes the sampled values of u^k, i.e., $u^k_{i,j} = u^k(i,j)$ for a suitable scaled image and τ is the time step size. The initial condition reads $u^0_i = f_i$ for all i.

5.4 Computational results

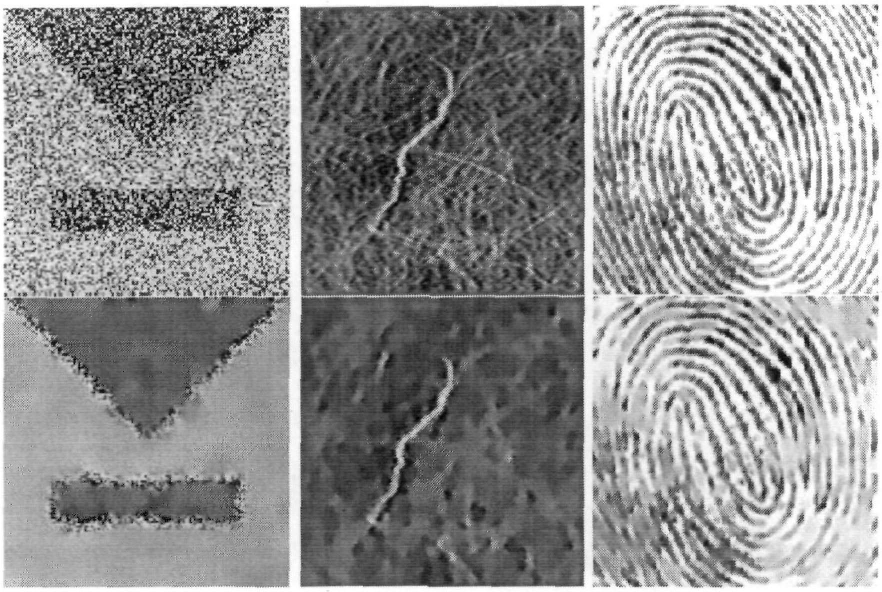

Figure 5.1: First row: Noisy images, Second row: Sharp operator based diffused images by taking parameters like $n = 4, \lambda = 3, \sigma = 3, t = 20$.

5.5 Parameter selection and observation

Nonlinear diffusion filtering contains several parameters that have to be specified in practical situations. The time t is an inherent parameter in each continuous diffusion process, so it has nothing to do with its discretization. The variance of evolving image decreases monotonically to zero and the contrast parameter λ is problem specific.

Our Sharp operator based model is mathematically sound and the results are similar to isotropic nonlinear diffusion filters of Weickert [72]. It has some drawbacks, it is not very sensitive at edges. Therefore, edge enhancing parameters can be introduced in order to preserve important features like edges.

Bibliography

[1] M.K. Ahmad and A.H. Siddiqi, *Image Classification and Comparative Study of Compression Techniques*, Sampling Theory in Signal and Image Processing - An International Journal, Σ Sampling Publishing, Leroy Street, Potsdam, NY - 13676, 1(2)(May 2002)155–180.

[2] R.A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[3] L. Alvarez, F. Guichard, P.L. Lions and J.M. Morel, *Axioms and Fundamental Equations of Image Processing*, Archive for Rational Mechanics and Analysis, 123(3)(1993)199–257.

[4] L. Alvarez, P.L. Lions and J.M. Morel, *Image Selective Smoothing and Edge Dection by Nonlinear Diffusion II*, SIAM Journal on Numerical Analysis, 29(3)(1992)845–866.

[5] F. Andreu et al., *Minimizing Total Variation Flow*, Differential and Integral Equations, 14(3)(2001)321–360.

[6] E. Arneodo, F. Argoul, E. Bary and J. Elezgaray, *Wavelet Transform of Fractals*, Y. Meyer (editor), (1992)286-352.

[7] PH. Benilan and H. Brezis, *Solutions Faibles D'évolution Dans les Espaces de Hilbert*, Ann. Inst. Fourier , 22(1972)311–329.

[8] C. Blatter, *Wavelets a Primer*, Massachussetts: AK Peters, 1998.

[9] H. Brezis, *Analyse Fonctionnelle, Théorie et Applications*, Masson, Paris, 1987.

[10] H. Brezis, *Operateurs Maximaux Monotones et Semi-groups de Contractions dans les Espaces de Hilbert*, Massachusetts Institute of Technology, Boston, MA, 1983.

[11] F. Catté, P.-L. Lions, J.M. Morel and T. Coll, *Image Selective Smoothing and Edge Detection by Nonlinear Diffusion*, SIAM Num. Analysis, 29(1992)182-193.

[12] P. Charbonnier et al., *Two Deterministic Half-quadratic Regularization Algoruthms for Computed Imaging,* In Proc. 1994 IEEE International Conference On Image Processing, V. 2, 168–172, Austin, TX, Nov. 1994, IEEE Computer Society Press.

[13] Q. Chen, and Z. Cheng, *A Study on Compactly Supported Orthogonal Vector-Valued Wavelets and Wavelet Packets,* Chaos, Solitons and Fractals, 31(2007)1024-1034.

[14] O. Christensen, *Frames, Riesz Bases and Discrete Gabor/Wavelet Expansions,* Bull. Amer. Math. Soc., 38(3)(2001)273–291.

[15] O. Christensen, *An Introduction to Frames and Riesz Bases,* Birkhäuser, Boston, 2003.

[16] C.K. Chui, *An Introduction to Wavelets,* Boston: Academic Press, Inc., 1992.

[17] C. K. Chui and Jain-ao Lian, *A study of Orthonormal Multi-wavelets,* Appl. Num. Maths., 20(1996)273-298.

[18] C.K. Chui and X. Shi, *Inequalities of Littlewood-Paley Type for Frame and Wavelets,* SIAM J. Math. Anal., 24(1)(1993)263–277.

[19] A. Cohen, I. Daubechies, and J.C. Feauveau, *Biorthogonal Bases of Compactly Supported Wavelets,* Comm. Pure Appl. Math., 45(1992)485–560, .

[20] R.R. Coifman and D.L. Donoho, *Translation Invariant De-noising,* in Wavelets and Statistics, ed. A. Antoniadis, Springer-Verlag, Heidelberg, 1995.

[21] R.R. Coifman and R. Rochberg, *Representation for holomorphic and harmonic functions in L_p,* Asterisque 77(1980)11–66.

[22] R. Crandall, *Project in Scientific Computation,* Springer-Verlag, New York, 211-212(1994)197–198.

[23] I. Daubechies, *Ten Lectures on Wavelets,* SIAM, Philadelphia, 1992.

[24] I. Daubechies, *The Wavelet Transform, Time-Frequency Localization and Signal Analysis,* IEEE Trans. Inform. theory, 36(5)(1990)961–1005.

[25] I. Daubechies, *Orthonormal Bases of Compactly Supported Wavelets,* Comm. Pure and Appl. Maths., 41(1988)909–996.

[26] I. Daubechies, A. Grossmann and Y. Meyer, *Painless Nonorthogonal Expansions,* J. Math. Phys., 27(1986)1271–1283.

[27] L. Debnath, ed. *Wavelet Transforms and their Applications*, Boston: Birkhäuser, 2002.

[28] D.L. Donoho, *Denoising by Soft-thresholding*, IEEE Trans. Inform. Theory, IT-41(1995)613–627.

[29] D.L. Donoho and I.M Johnstone, *Minimax Estimation by Wavelet Shrinkage*, Technical Report, Dept. of Statistics, Standferd University, 1992.

[30] G. Donovan, J. S. Geronimo, D. P. Hardin and P. R. Massopust, *Construction of Orthogonal Wavelets Using Fractal Interpolation Functions*, Preprint, 1994.

[31] G. Donovan, J. S. Geronimo and D. P. Hardin , *Intertwining Multiresolution Analysis and the Construction of Piecewise Polynomial Wavelets*, Preprint, 1994.

[32] R.J. Duffin and A.C. Schaeffer, *A Class of Non Harmonic Fourier Series*, Trans. Amer. Math. Soc., 72(1953)341–366.

[33] M. S. EI Naschie, *A Guide to the Mathematics of E-Infinity Contorian Space Time Theory*, Chaos, Solitons and Fractals, 25(2005)955-964.

[34] M. S. EI Naschie, *Hilbert, Fock and Contorian Spaces in the Quantum Two-Slit Gedanken Experiment*, Chaos, Solitons and Fractals, 27(2006)39-42.

[35] C. Fefferman and M. Stein, *H^p spaces of Several Variables*, Acta Mathematica, 129(1972)137–193.

[36] G. Gagneux, *Sur des Problèms Unilattéraux Dégénérés de la Théorie des Écoulements Diphasiques en Milieu Poreux*, These D'Etat, Universite de Besancon, 1982.

[37] H.Y. Gao, *Wavelet Shrinkage Denoised Using the Non-negative Garrote*, Journal of Comput. and Graphical Statistics, 7(4)(1998)469–488.

[38] H.Y. Gao and A.G.Bruce, *Wave Shrink with Firm Shrinkage*, Statistica Sinica, 7(1997)855–874.

[39] J. S. Geronimo, D. P. Hardin and P. R. Massopust, *Fractal Functions and Wavelet Expansions Based on Several Scaling Functions*, J.Approx.Theory, 1994.

[40] R.C. Gonzales and R.E. Wood, *Digital Image Processing*, Addison-Wesley Publishing Company, 1993.

[41] T. N. T. Goodman, S. L. Lee and W. S. Tang, *Wavelets in Wandering Subspaces*, Trans. Amer. Math. Soc., 338(1993)639–654.

[42] T. N. T. Goodman and S. L. Lee, *Wavelets of Multiplicity r*, Trans. Amer. Math. Soc., 342(1994)307–324.

[43] A. Grossmann, J. Morlet and T. Paul, *Transforms Associated to Square Integrable Group Representations I: General results*, J. Math. Phys. 26(1985)2473-2479, and
Transforms Associated to Square Integrable Group Representations II: Examples, Ann. Inst. Henri Poincare, Phys. Theor. 45(1986)293-309.

[44] G.H. Hardy and J.E. Littlewood, *A Maximal Theorem with Function-Theoretic Applications*, Acta Mathematica, 54(1930)81–116.

[45] C. Heil and D. Walnut, *Continuous and Discrete Wavelet Transforms*, SIAM Review, 31(1989)628-666.

[46] L. Hervé, *Multiresolution Analysis of Multiplicity d: Application to Dyadic Interpolation*, Appl. Computat. Harmonic Anal., 1(1994)199–315.

[47] B.B. Hubbard, *The World According to Wavelets*, Wellesley: A.K. Peters, 1996.

[48] T. Iijima, *Basic Theory on Normalization of Pattern (In Case of Typical One Dimensional Pattern*, Bulletin of the Electrotechnical Laboratory, 26(1962)368–388.

[49] F. John and L. Nirenberg, *On Functions of Bounded Mean Oscillation*, Comm. of Pure and Applied Maths, 15(1961)415–426.

[50] G. Kaiser, *A Friendly Guide to Wavelets*, Boston-Basel-Berlin: Birkhäuser, 1994.

[51] C. Kalisa and B. Torresani, *n-Dimensional Affine Weyl-Heisenberg Wavelets*, Ann. Inst. Henri Poincare, Phys. Theor., 59(2)(1993)201-236.

[52] S.L. Keeling and R. Stollberger, *Nonlinear Anisotropic Diffusion Filters for Wide Range Edge Sharpening*, Inverse Problem, 18(2002)175–190.

[53] J.J. Koenderink, *The Structure of Images* , Biological Cybernetics, 1984.

[54] J.L.Lions, *Controle Optimal de Systemes Gouvernés Pardes Équations aux Dérivées Partielles*, Dunod, Paris, 1968.

[55] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[56] S. Mallat, *A Theory of Multiresolution Signal Decomposition*, The Wavelet Representation, IEEE Trans. Patt. Anal. and Mach. Intell., 11(7)(1989)674–693.

[57] S. Mallat, *Multiresolution Approximation and Orthonormal Basis of $L^2(R)$*, Trans. Amer. Math. Soc., 315(1989)69–87.

[58] Y. Meyer, *Wavelets, Algorithms and Applications*, Philadelphia: SIAM, 1993.

[59] P. Mrázek and J. Weickert, *From two Dimensional Nonlinear Diffusion to Coupled Haar Wavelet Shrinkage*, J. Visual Comm. & Image Representation, 18(2007)162-175.

[60] M. Muszkieta, *Nonlinear Diffusion Filtering of Images Using Topological Gradient Approach to Edge Detection*, Ph.D. Thesis, University of Kaiserlautern, Germany, 2007.

[61] M. Nielsen, *Wavelets in Scientific Computing*, Ph.D. Dissertation, Department of Mathematical Modelling, Technical University of Denmark, 1998.

[62] P. Perona and J. Malik, *Scale Space and Edge Detection Using Anisotropic Diffusion*, IEEE Trans. Patt. Anal. Mac. Intell., 12(1990)629-639.

[63] W. Press et al., *Numerical Recipes in Fortran*, Cambridge University Press, New York, (1992)498–499.

[64] Z. Shang and X. Zhou, *Dual Generators for Weighted Irregular Wavelet Frames and Reconstruction Error*, Appl. Comp. Harm. Anal., 22(3)(2007)356–367.

[65] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley: Cambridge University Press, 1996.

[66] W. Sun and X. Zhou, *Irregular Wavelet/Gabor Frames*, Appl. Comp. Harm. Anal., 13(2002)63–76.

[67] L. Sun and Z. Cheng, *Construction of a Class of Compactly Supported Orthogonal Vector-Valued Wavelets*, Chaos, Solitons and Fractals, 34(2007)253-261.

[68] A. Torchinsky, *Real-Variable Methods in Harmonic Analysis*, Academic Press, Inc., 1986.

[69] B. Torresani, *Time-Frequency Representations: Wavelet Packets and Optimal Decomposition*, Ann. Inst. Henri Poincare, Phys. Theor., 56(2)(1992)215–234.

[70] B. Torresani, *Wavelets Associated with Representations of the Affine Weyl-Heisenberg Group*, J. Math. Phys., 32(5)(1991)1273-1279.

[71] M. Vetterli and C. Herley, *Wavelets and Filter Banks: Theory and Design*, IEEE Transactions on Signal Processing, 40(1992)2207-2232.

[72] J. Weickert, *Anisotropic Diffusion in Image Processing*, Stuttgart, Teubner, 1998.

[73] J. Weickert and B.M.H. Romeny, *Efficient and Reliable Schemes for Nonlinear Diffusion Filtering*, IEEE Transactions on Image Processing, 7(3)(March 1998)398-410.

[74] M. V. Wickerhauser, *Acoustic Signal Compression with Wavelet Packets*, In C.K. Chui (editor), (1991)679-700.

[75] M.V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters, 1994.

[76] D. Wirasael and S. Paolucci, *Adaptive Wavelet Method for Incompressible Flows in complex Domains*, Journal of Fluids Engineering, 127(2005)656-665.

[77] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*, London Mathematical Society, Students Texts 37, Cambridge University Press, 1997.

[78] S. Yang and X. Zhou, *A Note on Irregular Wavelet Frames*, Appl. Math. Comp., 187(2)(2007)1410–1416.

[79] X. Xia and B. W. Suter, *Vector Valued Wavelets and Vector Filter Banks*, IEEE Trans. Signal processing, 44(3)1996.

[80] R. Young, *An Introduction to Nonharmonic Fourier Series*, Academic Press, New York, 1980(revised first edition 2001).